

**STUDIES INTO THE OCCURRENCE OF  
 $\alpha$ -ONOCERIN IN RESTHARROW**

**BY**

**STEVEN PAUL HAYES  
BSc. (Hons), MSc.**

**Thesis submitted to the University of Nottingham for the  
degree of Doctor of Philosophy**

**November 2012**

**School of Biosciences  
Division of Plant and Crop Sciences  
The University of Nottingham, Sutton Bonington Campus  
Loughborough, Leicestershire, UK**

## ABSTRACT

With the increasing evidence of climate change in the coming decades, adaptive mechanisms present in nature may permit crop survival and growth on marginal or saline soils and is considered an important area of future research. Some subspecies of Restharrow; *O. repens subsp. maritima* and *O. reclinata* have developed the remarkable ability to colonise sand dunes, shingle beaches and cliff tops.  $\alpha$ -onocerin is a major component within the roots of Restharrow (*Ononis*) contributing up to 0.5% dry weight as described by Rowan and Dean (1972b). The ecological function of  $\alpha$ -onocerin is poorly understood, with suggestions that it has waterproofing properties, potentially inhibiting the flow of sodium chloride ions into root cells, or preventing desiccation in arid environments.

The fact that  $\alpha$ -onocerin (a secondary plant metabolite) biosynthesis has evolved a number of times in distantly related taxa; Club mosses, Ferns and Angiosperms, argues for a relatively simple mutation from non-producing antecedents. No direct research has been reported to have investigated the biosynthetic mechanism towards  $\alpha$ -onocerin synthesis via a squalene derived product originally characterised by Dean, and Rowan (1972a). A bi-cyclisation event of 2,3;22,23-dioxidosqualene by an oxidosqualene cyclase, may provide plants with an alternative mechanism for synthesising a range of triterpene diol products via  $\alpha$ -onocerin (Dean, and Rowan, 1972a). This mechanistic possibility presented the opportunity to investigate the biosynthesis of  $\alpha$ -onocerin using a multi-disciplinary approach.

This thesis presents supporting data, that  $\alpha$ -onocerin is a derivative of 2,3;22,23-dioxidosqualene via an oxidosqualene cyclase. Genetic markers were developed for *Ononis* versions of squalene cyclase, squalene epoxidase and a putative oxidosqualene cyclase. It was determined that squalene epoxidase; At4g37760 (SQE3), from *A. thaliana* showed the highest level of amino acid sequence conservation with the *O. spinosa* version. SQE3 is known to cyclise 2,3;22,23-dioxidosqualene (Rasbery *et al.*, 2007). Based on amino acid sequence alignments and predictive protein modelling a partial putative oxidosqualene cyclase isolated from *O. repens*

*subsp. maritima* is likely to be an *Ononis* version of  $\beta$ -amyrin synthase rather than a multifunctional oxidosqualene cyclase. Further functional characterisation studies are needed.

Methods were developed for analysing the transcriptome and metabolome in Restharrow which will aid future functional characterisation studies. Within *O. spinosa* root SQE3 was highly expressed. In contrast SQE3 was expressed at low levels in *O. spinosa* leaf and *O. pusilla* root and leaf. This data was supported by metabolomic profiling of five species of Restharrow; *O. spinosa*, *O. repens*, *O. repens subsp. maritima*, *O. pusilla* and *O. rotundifolia*. Triterpenes  $\alpha$ -onocerin and 2,3;22,23-dioxidosqualene were not present in *O. rotundifolia* and *O. pusilla*. Where 2,3;22,23-dioxidosqualene was present in plant extracts,  $\alpha$ -onocerin accumulation was also detected. *O. pusilla* and *O. spinosa* can be utilised for studying the occurrence of  $\alpha$ -onocerin within plants. The data presented in this thesis provides the necessary background information providing targets for functional expression studies of squalene epoxidases and oxidosqualene cyclases from Restharrow.

In summary the results in this thesis support the hypothesis proposed by Dean and Rowan (1971). There is good evidence to suggest the ability of Restharrow to cyclise  $\alpha$ -onocerin, may be dependant on the availability of 2,3;22,23-dioxidosqualene as the primary precursor. This was shown in development, tissue specific, ecotype and cell free enzyme analytical chemistry assays. There was little evidence to suggest a single specific oxidosqualene cyclisation event was primarily responsible for  $\alpha$ -onocerin biosynthesis. The work also presents evidence to suggest that differences in the squalene epoxidase sequence and transcription signals may affect the plants ability to cyclise  $\alpha$ -onocerin. This may have ecological implications and allow plants to adapt to their environment by providing an alternative route to biosynthesising membrane constituents via an alternative substrate specific mechanism.

## **ACKNOWLEDGEMENTS**

I express my gratitude to Dr. Sean Mayes for his extended patience, guidance, and unfaltering advice throughout my PhD. His encouragement and motivation has provided me with a fantastic start to a hopefully strong and successful career in science. I would also like to thank Dr. Peter D.G. Dean for his sound biochemical advice and support towards the completion of the project. Dr. Rob Linforth has played a fundamental role in guiding me through the use of analytical techniques in the laboratory and understanding principles behind metabolomic analysis.

Special thanks to Prof. Jerry Roberts for his support and thanks to Fiona Wilkinson, Julietta Marquez, Sue Golds, Emma Hooley, Chris Mills, Sheila Northover, Sue Flint and Mark Meacham and also colleagues Neil Graham, Dr. Shravani Basu, Dr. Chanate Malumpong, Dr. Kingsley J. Taah, Dr. Gracia V. Ribas and Dr. Peter Jones for their kind help and support.

I would like to thank Marilyn and Paul Hayes and Doreen Kilford, for their constant encouragement throughout the course of my PhD. In memory of my inspirational grandparents; Neil and Mary Hayes and Victor Kilford.

A special thanks to my beautiful wife Kirsty Hayes who has supported me morally, emotionally and financially and I am thankful for her unfaltering patience. My daughter Ffion Jo Hayes and son Isaac Steven Hayes have given me inspiration and a cause for continuing my quest towards completion during the writing stages of the thesis.

Last but not least thank you for the funding provided by Dr. Peter D.G. Dean through Cambio Ltd., the BBSRC and The University of Nottingham for providing the facilities necessary for the research.

## CONTENTS

<b>ABSTRACT</b>	<b>II</b>
<b>ACKNOWLEDGEMENTS</b>	<b>IV</b>
<b>TABLE OF CONTENTS</b>	<b>V</b>
<b>LIST OF TABLES</b>	<b>XII</b>
<b>LIST OF FIGURES</b>	<b>XVI</b>
<b>ABBREVIATIONS</b>	<b>XXVI</b>

## TABLE OF CONTENTS

<b>1</b>	<b>CHAPTER 1. INTRODUCTION</b>	<b>1</b>
1.1	RESTHARROW: A SPECIES OVERVIEW	1
1.2	SECONDARY METABOLITES	4
1.2.1	Introduction to secondary metabolites	4
1.2.2	Plant secondary metabolites	5
1.2.3	An introduction to $\alpha$ -onocerin	7
1.3	TRITERPENES	9
1.3.1	Introduction to triterpenes	9
1.3.2	The diversity of plant triterpenes	9
1.3.3	Triterpene biochemical pathway	11
1.4	OCCURRENCE AND SUGGESTED FUNCTION OF $\alpha$ -ONOCERIN	14
1.5	MEDICINAL USES OF $\alpha$ -ONOCERIN	20
1.6	BIODIVERSITY OF 2,3-OXIDOSQUALENE CYCLASES IN PLANTS	21
1.7	CYTOCHROME P450 HEMOPROTEINS	29
1.8	PROPOSED MECHANISTIC PATHWAYS FOR $\alpha$ -ONOCERIN BIOSYNTHESIS	31
1.9	POTENTIAL CYCLIC ROUTES TOWARDS $\alpha$ -ONOCERIN BIOSYNTHESIS	33
1.10	THE USE OF METABOLOMICS FOR BIOCHEMICAL CHARACTERISATION	33
1.11	METABOLOMICS IN PLANT BIOLOGY	34
1.12	AN INTRODUCTION TO THE TRANSCRIPTOME	35
1.13	ARABIDOPSIS AS A MODEL SPECIES	36
1.14	GENERAL PROJECT AIMS	36

<b>2</b>	<b>CHAPTER 2. MATERIALS AND METHODS</b>	<b>40</b>
2.1	PLANT METHODS	40
2.1.1	Plant materials	40
2.1.2	Seed sterilisation	40
2.1.3	Plant growth	40
2.2	NUCLEIC ACID METHODS	41
2.2.1	DNA extraction from plant leaf and root tissue	41
2.2.2	RNA extraction MOBIO kit	41
2.2.3	RNA extraction Qiagen RNeasy: high quality	42
2.2.4	RNA and DNA quantification	42
2.2.5	First strand cDNA synthesis	43
2.2.6	Reverse transcription	43
2.2.7	Polymerase Chain Reaction (PCR) methods	43
2.2.8	Rapid Amplification of cDNA ends (RACE) PCR	44
2.2.9	Nested PCR 2 <sup>nd</sup> round	45
2.2.10	Agarose Gel Electrophoresis	45
2.2.11	Agarose gel extraction of DNA molecules	45
2.2.12	PCR amplicon purification	46
2.2.13	DNA sequencing	46
2.2.14	Ligation of PCR products into pGEMT-easy vector	46
2.2.15	<i>E. coli</i> transformation	47
2.3	MICROARRAY METHODS	47
2.3.1	Oligonucleotide probe set identification	47
2.3.2	Oligonucleotide probe set cross hybridisation evaluation	48
2.3.3	Genomic DNA hybridisation; BioPrime® labeling	48
2.3.4	Hybridisation	48
2.3.5	Affymetrix GeneChip gDNA hybridisation procedure	48
2.3.6	Gene expression file creation	49
2.3.7	Genomic hybridisation intensity	49
2.3.8	RNA hybridisation	49
2.4	BIOINFORMATICS	49
2.4.1	Biochemical pathway gene characterisation	49
2.4.2	Use of genome databases for analysing the genome	50
2.5	ANALYTICAL CHEMICAL EXTRACTION METHODS	51
2.5.1	Plant sample preparation	51
2.5.2	Propan-2-ol extraction from plant tissue	51
2.5.3	Dichloromethane extraction from plant tissue	51
2.5.4	Ethanol extraction from plant tissue	52
2.5.5	Hexane/ diethyl ether extraction from plant tissue	52
2.6	GAS CHROMATOGRAPHY/ MASS SPECTROMETRY	53
2.6.1	Gas Chromatography using a Fisons 8000 series	53
2.6.2	Gas chromatography/mass spectrometry: Thermo Fisher DSQII	54
2.6.3	Gas chromatography/ mass spectrometry standardisation	55
2.6.4	Calibration of gas chromatogram peak area values	56
2.6.5	Mass spectral Selected Ion Remittance (SIR) detection	57
2.6.6	Quantification of non-saponifiable lipids	57

2.6.7	Experimental design temperature time extraction efficiency	58
2.7	EXPERIMENTS USING GC/MS SPECTROMETRY	59
2.7.1	Non-saponifiable lipid fraction of Restharrow throughout development	59
2.7.2	Non-saponifiable lipid fractions of Restharrow Ecotypes	59
2.7.3	Cell free plant extract -plant tissue preparation	60
2.7.4	Cell free extract: substrate inoculation	60
2.7.5	Extraction of cell free preparations: post incubation	60
<b>3</b>	<b>CHAPTER 3. THE USE OF PUBLICALLY AVAILABLE MICROARRAY DATABASES TO INVESTIGATE TRITERPENE SYNTHASE GENE EXPRESSION PATHWAYS A. THALIANA</b>	<b>62</b>
3.1	INTRODUCTION	62
3.1.1	General introduction to publically available genomic resources	62
3.1.2	An overview of micro array technologies	62
3.1.3	Affymetix GeneChip technology	63
3.1.4	Genevestigator as a resource for exploring microarray data	64
3.2	METHODS	65
3.3	RESULTS	65
3.3.1	AGI probes involved in triterpene biosynthesis in <i>Arabidopsis</i>	65
3.3.2	Gene specificity of probe sequences from squalene synthase, squalene epoxidase and oxidosqualene cyclases from <i>Arabidopsis thaliana</i> .	69
3.3.3	Comparative transcriptional profiling overview	71
3.3.4	Comparative transcriptional profiling squalene synthase	72
3.3.5	Comparative transcriptional profiling: squalene epoxidase	73
3.3.6	Comparative transcriptional profiling: oxidosqualene cyclases	75
3.3.7	Comparative transcriptional profiling: cycloartenol synthases	78
3.3.8	Metaprofile analysis: through plant development	79
3.3.9	Metaprofile analysis: organ specific comparative transcriptional profiling	82
3.4	DISCUSSION	83
3.4.1	Expression of genes involved in the triterpene pathway throughout development	83
3.4.2	Expression of genes involved in the triterpene pathway comparison between tissue types	85
3.4.3	Expression of genes involved in triterpene biosynthesis; a pathway overview	85

<b>4</b>	<b>CHAPTER 4. THE USE OF BIOINFORMATICS TO INVESTIGATE TRITERPENE BIOSYNTHESIS IN HIGHER PLANTS</b>	<b>87</b>
4.1	INTRODUCTION	87
4.1.1	General introduction to comparative bioinformatics	87
4.1.2	Hypothetical pathway model for $\alpha$ -onocerin biosynthesis	89
4.2	METHODS	90
4.2.1	Profiling triterpenoid biosynthesis using online tools	90
4.2.2	Use of online tools for genome mining of triterpenoid pathway	91
4.2.3	Analysis of protein sequence structure	92
4.3	RESULTS	92
4.3.1	Triterpene profiling: use of online databases	92
4.3.2	Comparative genome mining: Use of Aracyc combined metabolome PMN and genome tools	92
4.3.3	Arracyc: squalene synthase	93
4.3.4	Arracyc: squalene epoxidase	93
4.3.5	Arracyc: $\beta$ -amyrin synthase	94
4.3.6	Gene clustering of squalene epoxidases and $\beta$ -amyrin synthases	94
4.3.7	Development of degenerate molecular markers for genes involved in triterpene biosynthesis	95
4.3.8	Design of specific molecular markers from degenerate cDNA amplicons	98
4.3.8.1	Actin oligonucleotide marker	98
4.3.8.2	Squalene synthase Deg2 F2+R2	100
4.3.8.3	Squalene epoxidase oligonucleotide	101
4.3.8.4	Oxidosqualene cyclases	105
4.3.8.5	Putative $\beta$ -amyrin synthase	106
4.3.9	Phylogenetic relations between plant oxidosqualene cyclases	109
4.3.10	3D predictive modelling of oxidosqualene synthases	112
4.4	DISCUSSION	117
<b>5</b>	<b>CHAPTER 5. OCCURRENCE OF <math>\alpha</math>-ONOCERIN AND CYCLIC PRECURSORS IN RESTHARROW</b>	<b>123</b>
5.1	INTRODUCTION	123
5.1.1	General introduction	123
5.1.2	Triterpene biosynthesis in plants: squalene cyclisation	123
5.1.3	Cyclisation of 2,3-oxidosqualene	124
5.1.4	Natural products derivitised from 2,3;22,23-dioxidosqualene	126
5.1.5	Occurrence of triterpene diols	127
5.1.6	Post cyclic modifications of triterpenes	128
5.1.7	Proposed mechanistic pathways for $\alpha$ -onocerin biosynthesis	129
5.1.8	Summary	130
5.2	RESULTS: DERIVITISATION	131
5.2.1	Identification of internal standards from plant extracts	131



5.2.2	Mass spectral fragmentation patterns of alkanes	134
5.2.3	Percentage recovery experiment from NSL fractions	138
5.2.4	Optimisation and extraction efficiencies from plant extracts: temperature/time experiment	139
5.2.5	Mass spectral characteristics and identification of triterpenes	141
5.2.6	Identification of squalene using GC-EIMS	145
5.2.7	Identification of 2,3-oxidosqualene using GC-EIMS	146
5.2.8	Identification of 2,3;23,22-dioxidosqualene using GC- EIMS	146
5.2.9	Identification of $\alpha$ -onocerin ( $C_{30}H_{50}O_2$ ) using GC-EIMS	147
5.2.10	Identification of $\beta$ -amyrin ( $C_{30}H_{50}O$ ) using GC-EIMS	148
5.2.11	Identification of sitosterol ( $C_{30}H_{50}O$ ) using GC-EIMS	149
5.2.12	Identification of stigmasterol ( $C_{29}H_{48}O$ ), using GC-EIMS	150
5.2.13	Identification of campesterol ( $C_{28}H_{48}O$ ), using GC-EIMS	151
5.2.14	Hexane/diethyl ether extraction	152
5.3	RESULTS: EXPERIMENTAL	153
5.3.1	The occurrence of triterpenoids throughout plant development	153
5.3.2	Accumulation of triterpenes across Restharrow Ecotypes	154
5.3.3	Multi-species comparison of the triterpenoid metabolome	155
5.3.4	Triterpene biochemical profiles within individual species	155
5.3.5	In vivo studies of triterpene accumulation in cell preparation of Restharrow: cell free system For <i>Ononis</i> <i>spinosa</i> and cell free system for <i>O. pusilla</i>	161
5.4	DISCUSSION	162
5.4.1	Derivatisation of NSL triterpenoids	162
5.4.2	Developmental series	163
5.4.3	Squalene	163
5.4.4	2,3-oxidosqualene	164
5.4.5	2,3;22,23-dioxidosqualene	164
5.4.6	Stigmasterol and sitosterol	165
5.4.7	$\beta$ -amyrin	165
5.4.8	$\alpha$ -onocerin	165
5.4.9	Cell free enzyme reaction	167
<b>6</b>	<b>CHAPTER 6. X SPECIES TRANSCRIPTOMICS OF <i>O.SPINOSA</i> VS <i>O. PUSILLA</i>: GENES INVOLVED IN TRITERPENE BIOSYNTHESIS</b>	<b>168</b>
6.1	Transcriptional profil I INTRODUCTION	168
6.1.1	Transcriptional profiling	168
6.1.2	Cross species transcriptional profiling	170
6.2	RESULTS	172
6.2.1	Genomic DNA hybridisation and probe-selection	172
6.2.2	gDNA hybridisation intensities	172
6.2.3	Cross hybridisation analysis across four plant species	175
6.2.4	Squalene synthase hybridisation	176
6.2.5	Squalene epoxidase hybridisation	177

6.2.6	Oxidosqualene cyclase hybridisation	180
6.2.7	Putative triterpene synthase hybridisation	182
6.2.8	PM and MM analysis	184
6.3	TRANSCRIPTIONAL PROFILING	186
6.3.1	QC control figures and global expression difference between species	186
6.4	DIFFERENTIALLY EXPRESSED GENES	190
6.5	GENE EXPRESSION WITHIN THE TRITERPENE PATHWAY	195
6.5.1	Gene expression differences between two species of Restharrow	195
6.5.2	Gene expression squalene synthase	199
6.5.3	Gene expression squalene epoxidase	200
6.5.4	Gene expression oxidosqualene cyclase	201
6.6	DISCUSSION	201
<b>7</b>	<b>CHAPTER 7. GENERAL DISCUSSION</b>	<b>206</b>
7.1	ALPHA-ONOCERIN AND ITS IMPORTANCE	206
7.2	A POTENTIAL BIOSYNTHETIC MECHANISM TOWARDS $\alpha$ -ONOCERIN CYCLISATION	208
7.3	GENEVESTIGATOR CAN BE USED AS A TOOL TO IDENTIFY TRANSCRIPTIONAL CHANGES ACROSS BIOCHEMICAL PATHWAYS	209
7.4	PUBLICALLY AVAILABLE INFORMATION MAY OFFER AN INSIGHT INTO POTENTIAL TARGETS FOR $\alpha$ -ONOCERIN BIOSYNTHESIS INVESTIGATION	210
7.5	AMINO ACID SEQUENCE HOMOLOGY MAY SUGGEST FUNCTION OF PUTATIVE OXIDOSQUALENE CYCLASE ISOLATED FROM RESTHARROW	211
7.6	BIOCHEMICAL EVIDENCE TOWARDS $\alpha$ -ONOCERIN VIA A 2,3;22,23-DIOXIDOSQUALENE, CYCLISATION ROUTE	212
7.7	AFFYMETRIX ATH-1 CAN BE USED FOR CROSS SPECIES TRANSCRIPTOMICS OF RESTHARROW <i>O.</i> <i>PUSILLA</i> AND <i>O. SPINOSA</i>	214
7.8	PERFECT MATCH AND MISMATCH DATA OF INDIVIDUAL OLIGONUCLEOTIDES CAN BE USED TO IDENTIFY CROSS HYBRIDISATION	214
7.9	TRANSCRIPTIONAL EVIDENCE TOWARDS $\alpha$ - ONOCERIN CYCLISATION ROUTE VIA A 2,3;22,23- DIOXIDOSQUALENE CONTROLLED BY SQUALENE EPOXIDASE	217
7.10	SUMMARY: A PROPOSED MECHANISM FOR $\alpha$ -ONOCERIN BIOSYNTHESIS	218
7.11	FUTURE DIRECTIONS	219

<b>REFERENCES</b>	<b>221</b>
<b>APPENDIX</b>	<b>234</b>
APPENDIX I	MATERIALS AND METHODS USED IN THE THESIS 234
APPENDIX II	ARACYC TRITERPENE PATHWAY OUTPUT 237
APPENDIX III	GENE MODELS OF TRITERPENE BIOSYNTHESIS IN <i>A. THALIANA</i> 238
APPENDIX IV	CROSS HYBRIDISATION ASSESSMENT USING BLAST SEARCHES WITHIN AFFYMETRIX ONLINE TOOL RESOURCE 257
APPENDIX V	AMINO ACID SEQUENCE ALIGNMENT OF SQUALENE EPOXIDASES 276
APPENDIX VI	GENES AGI CODES AND PROBE SET ID NUMBERS INVOLVED IN THE PHYTOSTEROL BIOCHEMICAL PATHWAY IN <i>ARABIDOPSIS</i> 278
APPENDIX VII	PUTATIVE B-AMYRIN SYNTHASE NUCLEIC ACID SEQUENCE 279
APPENDIX VIII	GENE EXPRESSION PROFILES; META-PROFILE ANALYSIS TOOL FROM GENEVESTIGATOR 282
APPENDIX IX	GC/EIMS STANDARDS: GC, MS AND FRAGMENTATION IONS LISTED 283

## LIST OF TABLES

<b>Table 1.1</b>	A table showing the taxonomic classification, the common classification name, the taxonomic order of Restharrow, belonging to the family Fabaceae and Genus <i>Ononis</i> L.	1
<b>Table 1.2</b>	Classification of plant secondary metabolites; class, estimated number, functional role and commercial use. Data adapted from (Bravo, 1998).	6
<b>Table 1.3</b>	An outline of the molecular formula for $\alpha$ -onocerin, the CAS number and molecular weight of $\alpha$ -onocerin, as represented in the Chemical Abstracts Service (CAS), ( <a href="http://www.cas.org">www.cas.org</a> ).	8
<b>Table 1.4</b>	The table describes molecules related to $\alpha$ -onocerin, onoceroid derivatives and the plant species in which they are present. References are listed and adapted from Jacob <i>et al.</i> , 2004.	15
<b>Table 1.5</b>	The table displays the total $\alpha$ -onocerin content in milligrams per gram (mg/g) dry weight, isolated from root and aerial plant extracts of 12 <i>Ononis</i> species (Rowan, 1971).	18
<b>Table 1.6</b>	Characterised 2,3-oxidosqualene cyclase: cycloartenol synthase within plant species. The gene name and the primary plant metabolomic product are listed...Adapted from Herrera (1999) and Phillips <i>et al.</i> , (2006).	22
<b>Table 1.7</b>	Characterised 2,3-oxidosqualene cyclase: $\beta$ -amyrin synthase within plant species. The gene name and the primary plant metabolomic product are listed... Adapted from Herrera (1999) and Phillips <i>et al.</i> , (2006).	23
<b>Table 1.8</b>	Characterised 2,3-oxidosqualene cyclase: lupeol synthase within plant species. The gene name and the primary plant metabolomic product are listed... Adapted from Herrera (1999) and Phillips <i>et al.</i> , (2006).	25
<b>Table 1.9</b>	Characterised 2,3-oxidosqualene cyclase: mixed oxidosqualene cyclases within plant species. The gene name and the primary plant metabolomic product are listed...Adapted from Herrera (1999) and Phillips <i>et al.</i> , (2006).	26
<b>Table 2.1</b>	The table lists the PCR reaction mixture details, volume in $\mu$ l and concentration. NEB (New England Biolabs, Hitchin, UK).	44
<b>Table 2.2</b>	PCR thermocycler program parameters. The PCR step, time (minutes) and temperature ( $^{\circ}$ C) of each step and the number of cycles is listed.	44
<b>Table 2.3</b>	Online bioinformatics tools were used as listed in the table. The web link and the organisation source origin are also listed.	50

<b>Table 2.4</b>	The table below lists the injection parameters for the Thermo Scientific DSQ Trace GC Ultra. SSL (split/splitless) injector was subject to a gas pressure of 250 psi. The table lists the temperature, split flow rate, splitless time and the septum purge setting.	54
<b>Table 2.5</b>	The table below lists the operating mass spectrometry parameters including the; start time, detector gain, ion source charge, tune, scan rate and mass range, that were used to operate the GC-EIMS.	55
<b>Table 2.6</b>	A C <sub>28</sub> standard was used to calibrate peak area of 100% m/e saturation against concentration in µg/ml. The concentration of internal C <sub>28</sub> standard, volume of stock solution, amount of dichloromethane added are listed. The mean peak area and calculated peak are per µg/ml was also listed. 1µg/ml = a peak area of 1758 Mass lynx units.	56
<b>Table 2.7</b>	Two-factorial design using Design Expert programme 6.3, summarising experimental design parameters. Temperature in °C, and time in hours (hrs).	58
<b>Table 3.1</b>	The table lists secondary metabolites and genes involved in triterpenoid biosynthesis within <i>A. thaliana</i> . Listed are the abbreviations for the gene nomenclature and the Arabidopsis Gene Identification codes (AGI), (Auberg <i>et al.</i> 2002; Lange and Ghassemian 2003), used within GENEVESTIGATOR experiments.	66
<b>Table 3.2</b>	Oxidosqualene cyclase (At1g78950) probe ID 264138_at individual probe sequences 1-11 (NetAffx) and BLASTN (TAIR) sequence and score (E-value).	70
<b>Table 3.3</b>	The summary table outlined below provides an example of how the E-value related to the number of matched individual probe oligonucleotides (outlined in the Appendix IV).	71
<b>Table 4.1</b>	Three possible hypotheses for α-onocerin cyclisation within Restharrow. Hypothesis 1 suggests that the initial precursor is 2,3-oxidosqualene; Hypothesis 2 suggests the initial precursor is 2,3;22,23-dioxidosqualene; Hypothesis 3 suggests an alternative precursor is involved in α-onocerin synthesis.	90
<b>Table 4.2</b>	An organism comparison as a result of information output from triterpene pathway analysis using the Aracyc tool from the NASC database ( <a href="http://arabidopsis.info/">http://arabidopsis.info/</a> ).	93
<b>Table 4.3</b>	Degenerate primers designed to detect triterpene synthases in Restharrow...	97

<b>Table 4.4</b>	PCR amplified nucleotide sequence for a putative Actin from Restharrow ( <i>O. spinosa</i> ) cDNA using oligonucleotides; Actin F1 and Actin R1.	99
<b>Table 4.5</b>	PCR amplified nucleotide sequence for a putative squalene synthase from Restharrow ( <i>O. spinosa</i> ) cDNA using oligonucleotides; DEG-SQCY-F2 and DEG-SQCY-R2.	100
<b>Table 4.6</b>	PCR amplified nucleotide sequence for a putative squalene epoxidase from Restharrow ( <i>O. spinosa</i> ) cDNA using oligonucleotides DEG-SQE F2 and DEG-SQE R2.	102
<b>Table 4.7</b>	An Amino acid translation of putative squalene epoxidase (amplified PCR) nucleotide sequence from Restharrow ( <i>O. spinosa</i> ) cDNA using oligonucleotides; DEG-SQE F2 and DEG-SQE R2.	102
<b>Table 4.8</b>	The following species specific primers were designed from the sequenced products amplified using degenerate primers from <i>O. spinosa</i> cDNA template.	105
<b>Table 4.9</b>	Putative extended (from KLODA, 2005) $\beta$ -amyrin synthase amino acid sequence from Restharrow. Highlighted in yellow are the QW-motifs, MWCYCR is situated within the active site of the enzyme and SDCTAE ...	108
<b>Table 4.10</b>	A comparison between the active sites within oxidosqualene cyclases between closely related species within the legume family and unrelated <i>Arabidopsis</i> . Distantly related to legumes, resurrection plant, <i>Selaginella</i> , showed a highly conserved region within the active site.	111
<b>Table 5.1</b>	Possible outcomes for three hypotheses for $\alpha$ -onocerin cyclisation within Restharrow. Presence or absence of triterpenes and their precursors from within plant extracts.	130
<b>Table 5.2</b>	The physical properties of three long chain carbon molecules hexacosane, octacosane, triacontane belonging to the group; alkanes, used as internal standards for quantification of triterpenes.	132
<b>Table 5.3</b>	The table lists the name, isotope and atomic weight of small molecules which are constituents of triterpenes. The information was used for derivitisation methods.	135
<b>Table 5.4</b>	The table shows the percentage recovery of octacosane and $\alpha$ -onocerin, at a concentration of 1 mg/ml suspended in dichloromethane. Percentage recovery from dichloromethane and plant extract was measured.	138
<b>Table 5.5</b>	Peak areas of C <sub>28</sub> internal standard and $\alpha$ -onocerin from gas chromatograms produced as a result of the time/temperature experiment were calculated by	140

	Mass Lynx 3.2 software. Outlined in the table are the raw data, temperature (°C), time in hours, mean peak area of C <sub>28</sub> , mean peak area of $\alpha$ -onocerin and the calculated weight of $\alpha$ -onocerin (mg/g), dry weight.	
<b>Table 5.6</b>	The table provides a summary of retention times of triterpene metabolites and standards used for identification; analyte, retention time, molecular ion and molecular formula.	143
<b>Table 6.1</b>	A general summary of Cdf mask file statistics.	186
<b>Table 6.2</b>	A summary of correlation between species and tissue types. Cdf mask file PUSILLA100 (threshold 100) was used to measure the correlation between <i>O. pusilla</i> and <i>O. spinosa</i> root and leaf RNA hybridised onto ATH-1.	187
<b>Table 6.3</b>	A general summary of statistics. Cdf mask file spinosa 100 (threshold 100) was used to calculate expression values from <i>O. pusilla</i> and <i>O. spinosa</i> root and leaf RNA hybridised onto ATH-1.	187
<b>Table 6.4</b>	A summary of correlation between species and tissue types. Cdf mask file spinosa 100 (threshold 100) was used to assess correlation between <i>O. pusilla</i> and <i>O. spinosa</i> root and leaf RNA hybridised onto ATH-1.	187
<b>Table 7.1</b>	The oligonucleotide sequences where high intensity values occurred were; SQE1-6, SQE2-3, SQE3-8. The gene and the nucleotide sequence are listed.	215
<b>Table 7.2</b>	The amino acid sequences where high intensity values occurred were; SQE1, SQE2, SQE3. The gene and the amino acid sequence are listed. * represents sequence homology, represents amino acid differences.	216
<b>Table 7.3</b>	The amino acid sequences of active sites within SQE occurred were; SQE1, SQE2, SQE3, SQE4, SQE5, SQE6, SE11, SE12, ERG1 and ERG2.	216

## LIST OF FIGURES

<b>Figure 1.1</b>	Images of Restharrow: A) <i>O. repens subsp. maritima</i> , Southwold, UK. B) Representation of <i>Ononis spinosa</i> showing spines, leaf shape and the long tap root. C) <i>O. repens</i> , ascending stems, situated on a calcareous clay soil (Southwold, UK). D) <i>O. pusilla</i>	3
<b>Figure 1.2</b>	A diagrammatic representation of the stereochemical structure of $\alpha$ -onocerin.	7
<b>Figure 1.3</b>	Triterpene biosynthetic pathway: Acetyl Co-enzyme A, 3-hydroxy-3-methylglutaryl-coenzyme A, mevalonic acid (MVA), isopentenyl diphosphate (IPP), dimethylallyl diphosphate (DMAPP), farnesyl diphosphate (FPP), cycloartenol synthase (CYS), $\beta$ -amyrin synthase ( $\beta$ AS), lupeol synthase (LUS). (Suzuki <i>et al.</i> , 2002).	11
<b>Figure 1.4</b>	Schematic diagram showing the stages involved in the biosynthesis of 2,3-oxidosqualene.	13
<b>Figure 1.5</b>	Photographs of <i>O. repens subsp. maritima</i> in their natural environment (Hayes, 2008). A) showing the robust woody tap root, B) deep tap root growing on sandy shingle situated in a shingle habitat (Dunwich).	16
<b>Figure 1.6</b>	Rhizobial nodules on the roots of <i>Ononis spinosa</i> (Peter G D Dean, 2008).	19
<b>Figure 1.7</b>	An unrooted phylogram showing the relationship of triterpene synthases (cycloartenol synthase, $\beta$ -amyrin synthase, lupeol synthase, and multifunctional synthases) between species.	28
<b>Figure 1.8</b>	Action of cytochrome P450 catalytic cycle. Six stages in the catalytic cycle; $\text{Fe}^{2+}$ =oxidized cytochrome P450, $\text{Fe}^{2+}$ = reduced cytochrome P450, b5 = cytochrome b5, RH = substrate, ROH= product and $e^-$ = electron (adapted from Meunier <i>et al.</i> , 2004).	30
<b>Figure 1.9</b>	Proposed mechanistic pathway of triterpene biosynthesis. 2,3;22,23-dioxidosqualene cyclises initially to form a pre-onocerin intermediate.	33
<b>Figure 1.10</b>	Three hypothesis suggest alternatives for $\alpha$ -onocerin biosynthesis. Route 1; a direct cyclisation of 2,3-oxidosqualene, route 2) Cyclisation of a 2,3;22,23-dioxidosqualene, route 3) An alternative pathway.	37
<b>Figure 2.1</b>	The temperature and time settings of the gas chromatography oven were used for GC-EIMS derivitisation. The temperature gradient increased from 200°C to 315°C within 10 minutes. A temperature of 315°C was held for a further 8 minutes.	53



<b>Figure 2.2</b>	A calibration curve of internal standard octacosane, showing the concentration $\mu\text{g/ml}$ to peak area (Mass lynx units) ratio. The molecular ion (m/e) used for quantification of the octacosane was 97, which produced a 100% saturation of the mass spectrum.	57
<b>Figure 2.3</b>	A comparison of peak areas between hexacosane (RT 7.60 minutes) and $\alpha$ -onocerin RT 13.59 minutes), used for quantification.	58
<b>Figure 3.1</b>	A summary of the stages and genes involved in triterpene biosynthesis in <i>A. thaliana</i> deduced from Genevestigator, Plexdb and TAIR. Squalene synthase cyclising squalene At4g34640, At4g34650....legend continued...	68
		...69
<b>Figure 3.2</b>	The mean expression from all plant tissues combined; squalene synthetases (SQS), At4g24640 and At4g34650, were analysed at different developmental stages. The data was generated using high quality arrays, from within the Meta-profile analysis tool of GENEVESTIGATOR version 3 ( <a href="http://www.genevestigator.com">www.genevestigator.com</a> ).	73
<b>Figure 3.3</b>	The mean expression of all plant tissues combined from six squalene epoxidases (SQE): (At1g58440), At2g22830, At4g37760, At5g24140, At5g24150 and At5g24160 at different developmental stages was analysed.	74
<b>Figure 3.4</b>	The mean expression value from all plant tissues combined of oxidosqualene cyclases; At1g66960, At1g78950, At1g78960 and At1g78970 at different developmental stages was analysed.	76
<b>Figure 3.5</b>	The mean expression value from all plant tissues combined of six AtPEN oxidosqualene cyclases: At1g78500, At4g15370, At5g36150, At5g42600, At5g48010 at different developmental stages was analysed.	77
<b>Figure 3.6</b>	The mean expression value from all plant tissues combined of two cycloartenol synthases: as before At2g07050 and At3g45130 at different developmental stages was analysed.	79
<b>Figure 3.7</b>	A heat map generated by GENEVESTIGATOR: Meta profile tool showing a comparison of the mean expression value of individual genes within the triterpene pathway throughout development ( <a href="https://www.genevestigator.ethz.ch">https://www.genevestigator.ethz.ch</a> ).	80
<b>Figure 3.8</b>	Tissue specific gene expression of triterpene synthases displayed as a heat map. A comparison between general tissue groups; inflorescence, rosette and roots...	82

<b>Figure 4.1</b>	A flow diagram outlining the process of bioinformatics whilst investigating the suggested gene function. (Neil Graham, unpublished, <a href="http://plantsci.arabidopsis.info/pg/bio.html">http://plantsci.arabidopsis.info/pg/bio.html</a> ).	88
<b>Figure 4.2</b>	Chromosomal viewer output from Aracyc ( <a href="http://www.plantcyc.org">http://www.plantcyc.org</a> ), showing the chromosomal location of each of the gene homologues within the triterpene pathway within <i>A. thaliana</i>	95
<b>Figure 4.3</b>	A screen image from At1g78960 (multi-functional $\beta$ -amyrin synthase from <i>A. thaliana</i> ), translated nucleic acid sequence, is given as a central example of a TBLASTX, using the NCBI interface to search the Genbank database.	96
<b>Figure 4.4</b>	RT-PCR using degenerate primers designed from consensus sequences of <i>M. trunculata</i> , <i>P. sativum</i> and <i>A. thaliana</i> . Actin used as a PCR +ve control. Two sets of oligonucleotides were designed per gene. Squalene synthase (SQS F1-R1 and F2-R2), squalene epoxidase (SQE F1-R1 and F2-R2), $\beta$ -amyrin synthase (BAS F1-R1 and F2-R2)...	98
<b>Figure 4.5</b>	The results from a BLASTN search of the NCBI database, using a putative actin sequence amplified from Restharrow...	99
<b>Figure 4.6</b>	The results from a BLASTN search of the NCBI database, using a putative squalene synthase (Deg4) sequence amplified from Restharrow ( <i>O. spinosa</i> ).	101
<b>Figure 4.7</b>	A tBLASTx amino acid alignment between putative squalene epoxidase from Restharrow (Query SQE) and squalene epoxidase from <i>Medicago truncatula</i> (4g092640) and <i>Arabidopsis thaliana</i> At3g7760. 94% and 91% homology respectively...	103
<b>Figure 4.8</b>	RT-PCR amplified sequence using primers designed to a squalene epoxidase consensus was BLAST searched against the <i>A. thaliana</i> genome database.	103
<b>Figure 4.9</b>	A putative squalene epoxidase (ono) from <i>Ononis spinosa</i> , SQE1 (At1g58440), SQE2 (At2g22830) and SQE3 (At4g37760), translated amino acid sequences were aligned using Clustal W2 ( <a href="http://www.EBI.ac.uk">http://www.EBI.ac.uk</a> ).	104
<b>Figure 4.10</b>	An agarose gel electrophoresis image showing RT-PCR amplicons from oligonucleotides designed from <i>Ononis</i> genome sequence, deduced by degenerate PCR amplification.	106
<b>Figure 4.11</b>	A putative $\beta$ -amyrin synthase-like gene (Sean Mayes (unpublished), 2004), was subjected to an NCBI database BLASTN search.	107
<b>Figure 4.12</b>	NCBI BLASTP (Protein BLAST) of a putative partial $\beta$ -amyrin synthase-like gene (Kloda, 2005), was carried out to confirm whether the sequence was a putative $\beta$ -amyrin synthase...	109

<b>Figure 4.13</b>	Phylogram showing the relationship between $\beta$ -amyrin synthase amino acid sequences retrieved from a protein Basic Logical Alignment (BLASTP) from the NCBI database of a putative oxidosqualene cyclase (met13 protein) <i>Ononis spinosa</i> .	110
<b>Figure 4.14</b>	A phylogram showing the relationship between known oxidosqualene cyclases and the putative oxidosqualene cyclase from <i>Ononis</i> .	112
<b>Figure 4.15</b>	3D predictive modelling of an <i>Ononis</i> putative oxidosqualene cyclase (unpublished data, Mayes 2004). Output from Swiss Prot-automatic output viewer...	114
<b>Figure 4.16</b>	Secondary protein structure of (A) $\beta$ -amyrin synthase ( <i>P. sativum</i> ; BAA97558)...	115
<b>Figure 4.17</b>	Secondary protein structure of (C) $\beta$ -amyrin synthase ( <i>P. sativum</i> , BAA97558) overlaid with a putative oxidosqualene cyclase ( <i>O. spinosa</i> )...	115
<b>Figure 4.18</b>	Structural predictive 3D protein models of a putative oxidosqualene cyclase ( <i>O. spinosa</i> ) a $\beta$ -amyrin synthase ( <i>P. sativum</i> ; PSY, BAA97558), and a multifunctional oxidosqualene cyclase <i>P. sativum</i> ; PSM, BAA97559) modelled using Expasy SWISSprot modelling software. Major structural differences are indicated between the three protein models.	116
<b>Figure 5.1</b>	The triangular carbon-carbon-oxygen based terminal formed by the epoxide group affords a reactive intermediate when attached to squalene.	124
<b>Figure 5.2</b>	A representation of triterpene biosynthesis as it occurs in <i>A. thaliana</i> .	125
<b>Figure 5.3</b>	A representation of the conversion of 2,3;22,23 -dioxidosqualene to resinatenol oxide by a direct cyclisation event.	127
<b>Figure 5.4</b>	Stereo-chemical structure of $\beta$ -seco amyrin and $\alpha$ -seco amyrin.	128
<b>Figure 5.5</b>	The cyclisation of alpha and beta seco-amyrins from a 2,3-oxidosqualene derivative.	129
<b>Figure 5.6</b>	Gas Chromatogram derived from a Fisons 4,000 Gas Chromatography Mass Spectrometer. Internal standard: hexacosane, (C <sub>26</sub> H <sub>54</sub> ).	132
<b>Figure 5.7</b>	A Gas Chromatogram was derived from a Fisons 4,000 Gas Chromatography Mass Spectrometer. Internal standard octacosane (C <sub>28</sub> H <sub>58</sub> ).	133
<b>Figure 5.8</b>	An example of a gas chromatogram electron ionisation mass spectrum. GC-EIMS from within the internal peak area of an internal standard of triacontane, retention time 9.04 minutes (figure 5.7).	133
<b>Figure 5.9</b>	The figure represents an example of the carbon backbone configuration of alkanes; hexacosane, octacosane and triacontane.	134

<b>Figure 5.10</b>	A representative example of fragmentation patterns typically formed from a carbon (C <sub>8</sub> ) chain.	135
<b>Figure 5.11</b>	The total peak area of octacosane isolated by gas chromatography, was measured using Mass lynx software for calibration used for quantification.	136
<b>Figure 5.12</b>	The Total peak area values calculated from measuring the internal peak height of the most abundant m/e on the gas chromatogram, were used for calibration of known alkane standards.	136
<b>Figure 5.13</b>	A calibration curve showing the peak area originating from TIC m/e 394 against concentration of octacosane, used for calculating concentration of triterpenes.	137
<b>Figure 5.14</b>	A representative example of a Gas Chromatogram showing C <sub>28</sub> Alkane (RT 6.44 minutes) and α-onocerin (RT 13.82 minutes), from an ethanolic/potassium hydroxide extract of <i>O. pusilla</i> root tissue.	139
<b>Figure 5.15</b>	Design Expert interaction graph showing no significant effect of time (minutes) or temperature (°C) parameters on the extraction efficiency.	141
<b>Figure 5.16</b>	Gas Chromatography Electron Ionisation Mass Spectrometry (GC-EIMS), Total Ion Chromatogram (TIC) showing the retention time of external standards.	142
<b>Figure 5.17</b>	A gas chromatogram-electron ionisation mass spectrum of squalene (RT 11.79 minutes) from a Restharrow NSL plant root extract.	145
<b>Figure 5.18</b>	A gas chromatogram-electron ionisation mass spectrum of 2,3-oxidosqualene (squalene mono-epoxide), (RT 8.31 minutes), from a Restharrow NSL plant root extract.	146
<b>Figure 5.19</b>	A gas chromatogram-electron ionisation mass spectrum (RT 9.37 minutes) of 2,3;22,23-dioxidosqualene (squalene bis-epoxide), from a Restharrow NSL plant root extract.	147
<b>Figure 5.20</b>	A mass spectrum, of an analyte taken from a gas chromatogram retention time 11.20 minutes.	148
<b>Figure 5.21</b>	Mass Spectrum of β-amyrin (RT 10.41 minutes), from chromatogram figure 5.16. 100% saturation for molecular ion 218.	149
<b>Figure 5.22</b>	Gas Chromatography Electron Ionisation Mass spectrum of sitosterol showing relative abundance of ions and their molecular weight.	150
<b>Figure 5.23</b>	A GC-EIS mass spectrum from retention time of 9.90 minutes stigmasterol.	150

<b>Figure 5.24</b>	A gas chromatogram electron ionisation mass spectrum of campesterol as isolated from a NSL extract from Restharrow. Retention time of 9.71 minutes, figure 5.25.	151
<b>Figure 5.25</b>	A hexane/diethyl ether fraction from <i>O. spinosa</i> spiked sample with squalene, squalene mon-epoxide (2,3-oxidosqualene), squalene bis-epoxide (2,3;22;23-dioxidosqualene).	152
<b>Figure 5.26</b>	Two gas chromatograms showing NSL hexane/diethyl ether extracts from <i>A. thaliana</i> (top) and <i>O. repens maritima</i> (bottom).	152
<b>Figure 5.27</b>	Scatter plot showing metabolite accumulation (mg/g) in root tissue during development in <i>O. spinosa</i> .	153
<b>Figure 5.28</b>	A scatter plot showing metabolite accumulation; stigmasterol, sitosterol and $\beta$ -amyrin (mg/g dry weight) in <i>O. pusilla</i> root throughout development (weeks).	153
<b>Figure 5.29</b>	A total ion gas chromatogram from a NSL extract of <i>Ononis spinosa</i> root.	154
<b>Figure 5.30</b>	Total ion chromatogram showing relative abundance and retention time (minutes) from non-saponifiable lipid root extracts; biological triplicates from seven different plant species.	156
<b>Figure 5.31</b>	Total ion chromatogram showing relative abundance and retention time (minutes) of NSL leaf extracts of biological triplicates from seven different plant species.	157
<b>Figure 5.32</b>	Total ion chromatogram showing relative abundance and retention time (minutes) of NSL root and leaf extracts of biological triplicates from seven different plant species.	158
<b>Figure 5.33</b>	Histogram showing, bis-epoxide (2,3;22,23-dioxidosqualene), sitosterol, $\beta$ -amyrin, mono-epoxide (2,3-oxidosqualene) and $\alpha$ -onocerin quantities in milligrams per gram dry weight (mg/g d/w).	159
<b>Figure 5.34</b>	Histogram showing stigmasterol, bis-epoxide (2,3;22,23-dioxidosqualene), sitosterol, $\beta$ -amyrin, mono-epoxide (2,3-oxidosqualene) and $\alpha$ -onocerin quantities milligrams per gram dry weight (mg/g d/w).	160
<b>Figure 5.35</b>	Cell free enzyme incubation experiment. (Left) Expected outcome of substrates being passed through an incubated enzyme substrate complex. (middle) Peak area of products isolated from the enzyme substrate complex.	161
<b>Figure 5.36</b>	Cell free enzyme incubation experiment. Total peak area all treatments, boiled, incubated, no-substrate, no incubation, no boil or incubation controls. Substrates added	162

<b>Figure 6.1</b>	The central dogma of eukaryotic regulatory mechanism (Crick, 1970). The transcription from the DNA exon location, to RNA followed by the translation to protein.	168
<b>Figure 6.2</b>	Number of probe sets and probe pairs from the ATH1 – 121501 GeneChip array used to study the transcriptome from <i>O. pusilla</i> .	173
<b>Figure 6.3</b>	Number of probe sets and probe pairs from the ATH1 – 121501 GeneChip array used to study the transcriptome from <i>O. spinosa</i> .	174
<b>Figure 6.4</b>	Number of probe sets and probe pairs from the ATH1 – 121501 GeneChip array used to study the transcriptome from <i>A. thaliana</i> .	175
<b>Figure 6.5</b>	Normalised signal value from gDNA hybridisation on ATH-1. 253206_at, squalene synthase (At4g34640), (PM) hybridisation of four species compared; <i>A. thaliana</i> , <i>O. pusilla</i> , <i>O. spinosa</i> and <i>M. truncatula</i> .	176
<b>Figure 6.6</b>	Normalised signal value from gDNA hybridisation on ATH-1. 253254_at, squalene synthase (At4g34650), (PM) hybridisation of four species compared; <i>A. thaliana</i> , <i>O. pusilla</i> , <i>O. spinosa</i> and <i>M. truncatula</i> .	177
<b>Figure 6.7</b>	Normalised signal value from gDNA hybridisation on ATH-1. 245809_at, squalene epoxidase (At1g58440). (PM) hybridisation of four species compared; <i>A. thaliana</i> , <i>O. pusilla</i> , <i>O. spinosa</i> and <i>M. truncatula</i> .	177
<b>Figure 6.8</b>	Normalised signal value from gDNA hybridisation on ATH-1. 245809_at, squalene epoxidase (At1g58440). (MM) hybridisation of four species compared; <i>A. thaliana</i> , <i>O. pusilla</i> , <i>O. spinosa</i> and <i>M. truncatula</i> .	178
<b>Figure 6.9</b>	Normalised signal value from gDNA hybridisation on ATH-1. 266831_at, squalene epoxidase (At2g22830), (PM) hybridisation of four species compared; <i>A. thaliana</i> , <i>O. pusilla</i> , <i>O. spinosa</i> and <i>M. truncatula</i> .	178
<b>Figure 6.10</b>	Normalised signal value from gDNA hybridisation on ATH-1. 266831_at, squalene epoxidase (At2g22830).	178
<b>Figure 6.11</b>	Normalised signal value from gDNA hybridisation on ATH-1. 253039_at, squalene epoxidase, (At4g37760).	179
<b>Figure 6.12</b>	Normalised signal value from gDNA hybridisation on ATH-1. 253039_at, squalene epoxidase, (At4g37760),	179

<b>Figure 6.13</b>	Normalised signal value from gDNA hybridisation on ATH-1. 249773_at, squalene epoxidase (At5g24140), (PM) hybridisation of four species compared; <i>A. thaliana</i> , <i>O. pusilla</i> , <i>O. spinosa</i> and <i>M. truncatula</i> .	179
<b>Figure 6.14</b>	Normalised signal value from gDNA hybridisation on ATH-1. 249774_at, squalene epoxidase (At5g24150), (PM) hybridisation of four species compared; <i>A. thaliana</i> , <i>O. pusilla</i> , <i>O. spinosa</i> and <i>M. truncatula</i> .	180
<b>Figure 6.15</b>	Normalised signal value from gDNA hybridisation on ATH-1. 249775_at, squalene epoxidase (At5g24160).	180
<b>Figure 6.16</b>	Normalised signal value from gDNA hybridisation on ATH-1. 255912_at, Lupeol synthase (At1g66960).	180
<b>Figure 6.17</b>	Normalised signal value from gDNA hybridisation on ATH-1. 264138_at, $\beta$ -amyrin synthase (At1g78950).	181
<b>Figure 6.18</b>	Normalised signal value from gDNA hybridisation on ATH-1. 264137_at, $\beta$ -amyrin synthase (At1g78960).	181
<b>Figure 6.19</b>	Normalised signal value from gDNA hybridisation on ATH-1. 264100_at, Lupeol synthase (At1g78970).	181
<b>Figure 6.20</b>	Normalised signal value from gDNA hybridisation on ATH-1. 263123_at, Triterpene synthase (At1g78500).	182
<b>Figure 6.21</b>	Normalised signal value from gDNA hybridisation on ATH-1. 245258_at, Triterpene synthase (At4g15340).	182
<b>Figure 6.22</b>	Normalised signal value from gDNA hybridisation on ATH-1. 245553_at, Triterpene synthase (At4g15370).	182
<b>Figure 6.23</b>	Normalised signal value from gDNA hybridisation on ATH-1. 249687_at, Triterpene synthase (At5g36150).	183
<b>Figure 6.24</b>	Normalised signal value from gDNA hybridisation on ATH-1. 249205_at, Triterpene synthase (At5g42600).	183
<b>Figure 6.25</b>	Normalised signal value from gDNA hybridisation on ATH-1. 248729_at, Triterpene synthase (At5g48010).	183
<b>Figure 6.26</b>	Normalised mean hybridisation signal for <i>O. spinosa</i> and <i>O. pusilla</i> PM and MM oligonucleotide probe sets. Genes listed in table 3.1	185
<b>Figure 6.27</b>	Hybridisation controls for <i>O. pusilla</i> and <i>O. spinosa</i> root and leaf tissue, using mask files; A, <i>pusilla</i> 100 and B, <i>spinosa</i> 100.	188

<b>Figure 6.28</b>	Box whisper plot of the transcriptional profile using the mask pusilla100 (100 threshold), RMA normalised.	188
<b>Figure 6.29</b>	Box whisper plot of the transcriptional profile using the mask spinosa100 (100 threshold), RMA normalised.	189
<b>Figure 6.30</b>	A Histogram of the transcriptional profile using the mask pusilla100 (100 threshold), RMA normalised, showing normal distribution of signal intensities between <i>O. pusilla</i> root, <i>O. pusilla</i> leaf, <i>O. spinosa</i> root and <i>O. spinosa</i> leaf.	189
<b>Figure 6.31</b>	A Histogram of the transcriptional profile using the mask spinosa100 (100 threshold), RMA normalised, showing normal distribution of signal intensities between <i>O. pusilla</i> root, <i>O. pusilla</i> leaf, <i>O. spinosa</i> root and <i>O. spinosa</i> leaf.	190
<b>Figure 6.32</b>	An overview of normalised expression trajectories of differentially expressed genes between <i>O. pusilla</i> Root (OP-Rootp100), <i>O. pusilla</i> Leaf (OP-Leafp100), <i>O. spinosa</i> root (OS-Root p100) and <i>O. spinosa</i> Leaf (OS-leafp100).	191
<b>Figure 6.33</b>	An overview of normalised expression trajectories of differentially expressed genes between <i>O. pusilla</i> Root (OP-Rootp100), <i>O. pusilla</i> Leaf (OP-Leafp100), <i>O. spinosa</i> root (OS-Root p100) and <i>O. spinosa</i> Leaf (OS-leafp100).	192
<b>Figure 6.34</b>	An overview of normalised expression trajectories of differentially expressed genes between <i>O. pusilla</i> Root (OP Root), and <i>O. spinosa</i> root (OS Root).	193
<b>Figure 6.35</b>	An overview of normalised expression trajectories of differentially expressed genes between <i>O. pusilla</i> Root (OP Root), and <i>O. spinosa</i> root (OS Root).	194
<b>Figure 6.36</b>	Overview of normalised signal expression value of genes involved in the triterpene biosynthetic pathway. Transcriptional profile between <i>O. pusilla</i> Root, <i>O. pusilla</i> leaf, <i>O. spinosa</i> root and <i>O. spinosa</i> leaf.	196
<b>Figure 6.37</b>	Normalised signal expression value of differentially expressed oxidosqualene cyclases involved in the triterpene biosynthetic pathway.	197
<b>Figure 6.38</b>	Normalised signal expression value of differentially expressed. Oxidosqualene cyclases involved in the triterpene biosynthetic pathway. Transcriptional profile between <i>O. pusilla</i> Root, <i>O. pusilla</i> leaf, <i>O. spinosa</i> root and <i>O. spinosa</i> leaf.	198
<b>Figure 6.39</b>	Normalised signal expression value of differentially expressed squalene synthases involved in the triterpene biosynthetic pathway.	199



<b>Figure 6.40</b>	Normalised signal expression value of differentially expressed squalene epoxidases involved in the triterpene biosynthetic pathway. Transcriptional profile between <i>O. pusilla</i> root, <i>O. pusilla</i> leaf, <i>O. spinosa</i> root and <i>O. spinosa</i> leaf.	200
<b>Figure 6.41</b>	Normalised signal expression value of differentially expressed oxidosqualene cyclases involved in the triterpene biosynthetic pathway. Transcriptional profile between <i>O. pusilla</i> Root, <i>O. pusilla</i> leaf, <i>O. spinosa</i> root and <i>O. spinosa</i> leaf.	201
<b>Figure 7.1</b>	Potential route towards $\alpha$ -onocerin biosynthesis via an oxidosqualne cyclase.	208
<b>Figure 7.2</b>	Triterpene biosynthetic pathway within the root of Restharrow <i>O. spinosa</i> , <i>O. repens</i> and <i>O. repens</i> , <i>subsp- maritima</i> .	219

## ABBREVIATIONS

AGI	Arabidopsis Genome Initiative
ATPEN	Arabidopsis thaliana pentacyclic triterpene synthase
BLAST	Basic Local Alignment Search Tool
BAS	$\beta$ -amyrin synthase
cDNA	complementary Deoxyribonucleic acid
CYC	Cycloartenol synthase
C <sub>x</sub>	Carbon
°C	Degrees Celsius
CAM	Crassulacean acid metabolism
DOS	2,3;22,23-dioxidosqualene
DNA	Deoxyribonucleic acid
DMAPP	Dimethylallyl diphosphate
dNTP	Deoxynucleotide tri phosphate
d/w	Dry weight
EBI	European Bioinformatics Institute
EIMS	Electron ionization mass spectrometry
EST	Expressed sequence tag
FPP	Farnesyl di-phosphate
FPS	Farnesyl diphosphate synthase
gDNA	genomic DNA
g	Grams
GC-MS	Gas chromatography – mass spectrometry
GEO	Gene Expression Omnibus
Glu	Glutamic acid
Gly	Glycine
g/mol	Grams per mol
H	Hydrogen
H <sub>2</sub> O	Water
HPLC	High performance liquid chromatography
IPP	Isopentenyl diphosphate
KEGG	Kyoto Encyclopedia of Genomes and Genes
KOH	Potassium hydroxide

LUP	Lupeol synthase
LSS	Lanosterol synthase
MEGA	Molecular Evolutionary Genetics Analysis
mg	Milligrams
mins	Minutes
MM	Mismatch
mRNA	messenger Ribonucleic acid
MRN	Maneral synthase
MS	Murashige and Skoog
m/z, m/e	Molecular ion
MVA	Mevalonic acid
NASC	Nottingham Arabidopsis Stock Centre, UK
NCBI	The National Center for Biotechnology Information
NIST	National Institute of Standards and Technology
NMR	Nuclear magnetic resonance
NSL	Non-saponifiable lipid
O	Oxygen
OH	Hydroxyl group
ORF	Open Reading Flame
OS	2,3-oxidosqualene
OSC	Oxidosqualene cyclase
PCR	Polymerase Chain Reaction
pLEXdb	Plant Expression Database
PM	Perfect Match
PNY	<i>Panax gensing</i> $\beta$ -amyrin synthase
PP	Diphosphate
PSPP	Pre-squalene pyrophosphate
Q-PCR	Quantitative Polymerase Chain Reaction
RACE	Rapid Amplification of cDNA Ends
R <sup>2</sup>	Regression line
RNA	Ribonucleic acid
RT	Retention time
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SQS	Squalene synthase

SQE	Squalene epoxidase
ssDNA	single-strand DNA
TAIR	The Arabidopsis Information Resource
tBLASTX	Translated Basic Local Alignment Search Tool
T-DNA	Transfer-Deoxyribo Nucleic Acid
THA1	Thalianol synthase 1
TIGR	The Institute for Genomic Research
TTPS	Triterpene synthase
UV	Ultraviolet
X-ray	Xerography

# CHAPTER 1. INTRODUCTION

## 1.1 RESTHARROW: A SPECIES OVERVIEW

The common name for plants belonging to the *Ononis* genus (Leguminosae) is Restharrow. In this group of perennials, 68 species have currently been described (Crespo, 1991). The largest concentration of Restharrow can be found in coastal areas in the Western Mediterranean region, with half of these species also found growing in North Africa. Restharrow has also been reported to be found as far East as Serbia (Jaric *et al.*, 2007). Restharrow generally flourishes on dry and calcareous soils; however they are not especially adapted to high humidity or extreme temperatures (Ivimey-Cook, 1968).

**Table 1.1** A table showing the taxonomic classification, the common classification name, the taxonomic order of Restharrow, belonging to the family Fabaceae and Genus *Ononis* L.

Taxonomic Classification	Taxonomic name	Common name classification
Kingdom	Plantae	Plants
Subkingdom	Tracheobionta	Vascular plants
Superdivision	Spermatophyte	Seed plants
Division	Magnoliophyta	Flowering plants
Class	Magnoliopsida	Dicotyledons
Subclass	Rosidae	Rosids
Order	Fabales	
Family	Fabaceae	Pea family
Genus	<i>Ononis</i> L.	Restharrow
Species	<i>Ononis spinosa</i> , <i>Ononis repens</i> , <i>Ononis pusilla</i> , <i>Ononis rotundifolia</i>	Spiny Restharrow Common Restharrow

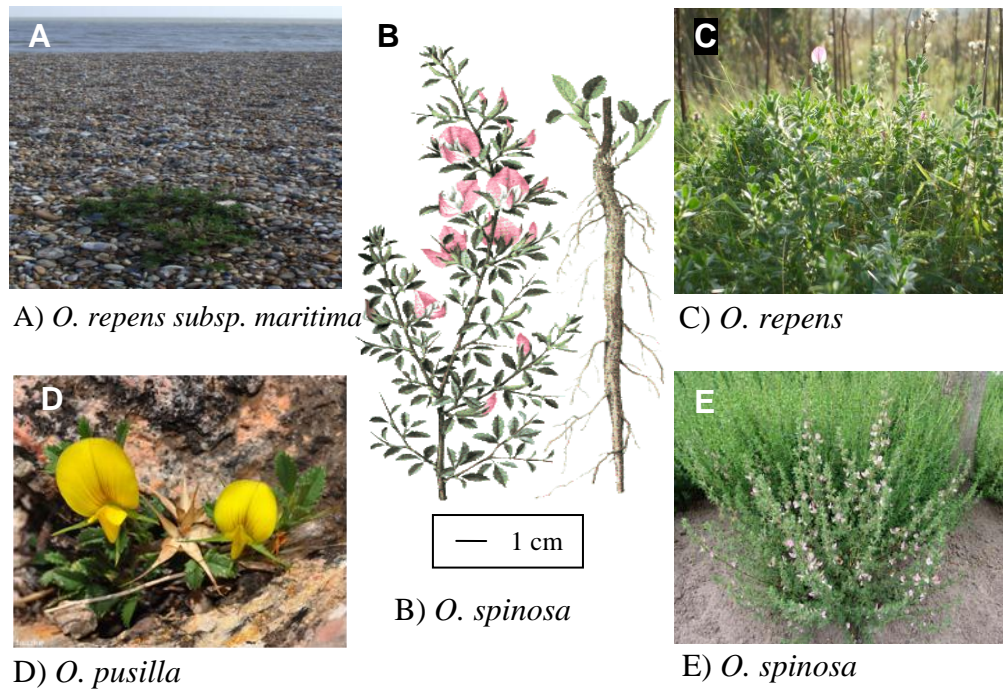
In the UK, *Ononis spinosa* and *Ononis repens* are the predominant species. *Ononis spinosa* occupies chalk, limestone and calcareous clay soils, whereas

*O. repens* can be found situated on well drained light soils, common grasslands and boulder-clays. *Ononis repens* subsp. *maritima* can be found occupying coastal sands, shingle beaches and cliff tops. What has been described as *Ononis spinosa* subsp. *intermedia* can be found in a few areas as stable populations. *Ononis reclinata* has been reported to occupy maritime habitats, although they persist only in a few coastal areas around Britain (Ivimey-Cook, 1968).

The taxonomy of Restharrow has become a long running debate as to whether *O. repens* and *O. spinosa* can interbreed to form an intermediate. There is also a question as to whether the differences in their morphology are simply a direct result of interactions with the surrounding environment. Current genetic data using principle coordinate analysis of microsatellite loci (Kloda *et al.*, 2008), supports the hypothesis that interbreeding between *O. repens* and *O. spinosa* does not occur regularly, although interbreeding between *O. repens* and *O. repens* *maritima* and also *O. spinosa* and *O. intermedia*, respectively, is more likely. This is not intrinsically suprising, as *O. spinosa* and *O. intermedia* are diploid species whereas *O. repens* and *O. repens* *maritima* are tetraploid species, so interbreeding would be expected to lead to a triploid block (Ivemy-Cook, 1968). One of the chemotaxonomic traits that has been utilised for Genera identification is the ability of Restharrow to produce the secondary metabolite  $\alpha$ -onocerin. In 1972 it was discovered that not all *Ononis* species produce the triterpenoid  $\alpha$ -onocerin (Rowan and Dean, 1972b).

Restharrow has characteristic morphological features such as a long tap root and are often described as rhizomatous perennials, depending upon the species and the habitat they occupy. Generally Restharrow has a low growth habit, varying from procumbent to ascendant (60 cm). Restharrow is often described as woody at their base with obtuse to emarginated leaves, growing in a loose formation. Generally the leaves are 3 times as long as their width with flowers protruding from terminal racemes (figure 1.1).

To date there have been very few ecological studies of Restharrow and their impact on the natural environment. Batanouny (2001) argues that Restharrow occupying coastal regions, are thought to be important in stabilizing sand dunes due to the long tap root system. Restharrow are commonly found as a pioneering species during the yellow and grey sand dune development stages.



**Figure 1.1** Images of Restharrow: A) *O. repens subsp. maritima*, Southwold, UK. B) Representation of *Ononis spinosa* showing spines, leaf shape and the long tap root. C) *O. repens*, ascending stems, situated on a calcareous clay soil (Southwold, UK). D) *O. pusilla* ([http://www.bioscripts.net/flora/index.php?spp=Ononis %20pusilla](http://www.bioscripts.net/flora/index.php?spp=Ononis%20pusilla)), E) *O. spinosa* occupying a sand dune habitat.

*O. pusilla* (figure 1.1, D) is a species of Restharrow which is common throughout the Mediterranean region of Southern Europe. The growth habit of *O. pusilla* is similar to that of *O. repens* and it can form substantial shrubs. *O. pusilla* can be found occupying rocky outcrops, stony ground and

open pine forest habitats. Cytogenetically *O. pusilla* has a chromosome number of  $2n = 2x = 30$ , as does *O. spinosa* which is also a diploid ( $2n = 2x = 30$ ). *O. repens* and *O. repens subsp. maritima* are tetraploids both having chromosome counts of  $2n=4x=60$  (Morisset, 1978).

The ability of Restharrow to occupy a wide variety of natural habitats, from saline saturated (halophytic) beaches, to semi-arid environments is thought to be a result of the ability of the plant to produce secondary metabolites (Gerd Liebezeit, 2007). One group of secondary metabolites found in abundance within Restharrow are the triterpenes, which are  $C_{30}$  compounds produced as part of the isoprenoid pathway.

## **1.2 SECONDARY METABOLITES**

### **1.2.1 Introduction to secondary metabolites**

Secondary metabolites are ‘organic compounds that are not directly involved in the normal growth, development or reproduction within an organism’ (Corey *et al.*, 1997). The presence of secondary metabolites may affect the ability of the plant to adapt to changes in biotic or environmental conditions. According to Karlovsky (2008), primary metabolites are the essential components responsible for normal cellular functionality including, cell division, growth and reproduction. Secondary metabolites offer potential ecological advantages to an organism such as defence against predators, parasites and diseases and may allow organisms to occupy ecological niches, which would otherwise be occupied by competitors.

Secondary metabolites occur across a diverse range of organisms including animals, plants, bacteria, fungi, and marine organisms. Secondary metabolites are widely utilised for commercial uses, some examples are; terpenes (flavour compounds), alkaloids (opiates) and glycosides (antifungal agents). Secondary metabolites are considered as important for a wide range of industries as they are used for pharmaceutical, household chemical, food technology and fuel commercial purposes (Zhong *et al.*, 2005). For this reason it is important to understand the origin, biological occurrence, distribution and biosynthesis of these compounds. Secondary metabolites



play an important part in the physiological makeup of plant cells and contribute towards processes which may allow plants to adapt to environmental habitats. As secondary metabolites are derived from the combination and modification of the products of primary metabolites, they can be divided into categories based on their biosynthetic origin. The main groups are; alkaloids, terpenoids, glycosides, phenols, phenazines, polyketides, fatty acids and peptides.

### **1.2.2 Plant secondary metabolites**

To date 30,000 biochemical compounds have been described as plant secondary metabolites (Zhong *et al.*, 2005). Gang *et al.*, (2001) report that not all plant groups produce secondary metabolites, and that secondary metabolites are often produced at specific time points according to plant requirements during the developmental stages. Plants have developed the remarkable ability to up-regulate secondary metabolites as a response to environmental stimuli which often leads to accumulation in specific tissues. (Argueso *et al.*, 2009). As a result of evolutionary processes, groups of secondary metabolites have evolved that are specific to particular taxonomic groups and often unique between species (Wink, 2003). Some secondary metabolites (e.g. caffeine), have evolved in taxonomically unrelated plant species as a result of convergent evolution (Bennet and Wallsgrove, 1994).

Triterpenoid saponins and phytosterols (groups of secondary metabolites) can be biosynthesised as a response to biotic or abiotic stimuli (Creelman and Mullet, 1995). The accumulation of secondary metabolites in plants is often localised to specific tissues which can subsequently be transported to tissues where they are required for functional roles. One example of this is the occurrence of wax compounds in root tissues that inhibit the flow of sodium chloride ions and water across the root membrane in mangroves occupying coastal regions (Bandaranyake, 2002; Barrett-Lennard, 2003).

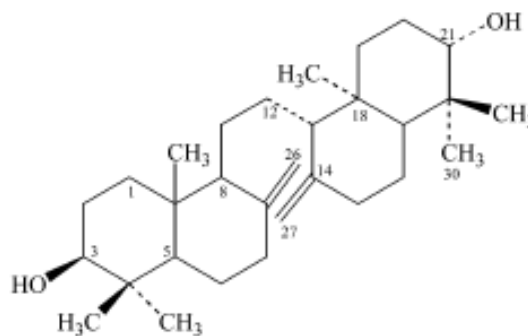
**Table 1.2** Classification of plant secondary metabolites; class, estimated number, functional role and commercial use. Data adapted from (Bravo, 1998).

Class of secondary metabolites	Estimated number of metabolites	Functional role within the plant	Commercial use
Terpenoids	~100,000	Plant defence Membrane sterols Phytohormones Saponins	Essential oils insecticides Medicines (e.g. taxol, artemisinin)
Alkaloids	~12,000	Plant defence (e.g. caffeine)	Insecticides poisons Morphine
Phenylpropanoids and phenolics	~8,000	Plant defence, Structure: - bark - lignins - flavanoids	Spices Flavours (e.g. vanillin)

Restharrow is known to accumulate large quantities of secondary metabolites within the root cortex (Rowan and Dean, 1972b). It is thought that secondary metabolites may provide a barrier that helps to prevent the movement of sodium chloride ions across the root membrane in saline environments (Spikett *et al.*, 1993), allowing the colonisation of coastal areas. Several studies of Restharrow have revealed a significant amount of  $\alpha$ -onocerin accumulates within the root.

### 1.2.3 An introduction to $\alpha$ -onocerin

The biosynthesis of the triterpenoid  $\alpha$ -onocerin is poorly understood at the molecular and biochemical levels. The chemical structure (figure 1.2), consists of four cyclic rings coupled by an incomplete cyclic isoprene unit in the central position. The molecule is symmetrical, which for a triterpene is unusual in nature.



**Figure 1.2** A diagrammatic representation of the stereochemical structure of  $\alpha$ -onocerin. OH indicates a hydroxyl group, CH<sub>3</sub> = methyl group. Numbers represent carbon positions. Double line indicates double bond.

The triterpenoid  $\alpha$ -onocerin was first extracted from the roots of *Ononis spinosa* by Hlasiwetz in 1855. Further investigation into the stereochemistry of  $\alpha$ -onocerin and its derivatives was continued over the next century by a number of biochemists; (von Hemmelmayr, 1907; Dieterle *et al.*, 1934; Zimmermann, 1938 and 1940).  $\alpha$ -Onocerin was the first squalenoid to be fully classified in the vegetable kingdom. Barton and Overton (1955) further characterised  $\alpha$ -onocerin and  $\gamma$ -onocerin (formed by acid treatment of  $\alpha$ -onocerin).  $\beta$ -onocerin was first prepared by Zimmermann (1938), and its inorganic synthesis was finally achieved by acid treatment of  $\alpha$ -onocerin diacetate (Schaffner *et al.*, 1956).

Theoretical derivatives;  $\gamma$ -onocerin and  $\beta$ -onocerin, have to date not been detected as natural plant products, as has been discussed by Boiteau *et al.*, (1962). The structures of both  $\alpha$ -onocerin and  $\beta$ -onocerin were elucidated using X-ray crystallography (Stork *et al.*, 1963, van Tamelen, 1968; Mi *et al.*, 2002, Pauli, 2000; Fröhlich and Pauli, 2000).

**Table 1.3** An outline of the molecular formula for  $\alpha$ -onocerin, the CAS number and molecular weight of  $\alpha$ -onocerin, as represented in the Chemical Abstracts Service (CAS), ([www.cas.org](http://www.cas.org)).

<b>CAS number</b>	511-01-3
<b>Molecular weight</b>	442 (M/e)
<b>Molecular formula</b>	$C_{50}H_{30}O_2$
<b>Synonyms</b>	(3 $\beta$ ,21 $\alpha$ )-8,14-secogammacera-8(26),14(27)-diene-3,21-diol
	5 $\beta$ ,5' $\beta$ -Ethylenebis[(2S,8 $\alpha\alpha$ )-1,1,4 $\alpha\beta$ -trimethyl-6-methylenedecalin-
	8,14-Secogammacera-8(26),14(27)-diene-3 $\beta$ ,21 $\alpha$ -diol

Studies into the occurrence and synthesis of  $\alpha$ -onocerin in *Restharrow*s by Rowan and Dean (1972b), contributed further towards an understanding of  $\alpha$ -onocerin synthesis (Rowan and Dean, 1972a). Rowan and Dean (1972a), determined that  $\alpha$ -onocerin is derived from the triterpenoid pathway and is a product of a squalene enzyme substrate complex. It is thought that  $\alpha$ -onocerin is thought to be synthesised by a process infrequently found in nature: a cyclisation event from both termini of a 2,3;22,23-dioxidosqualene (bis-epoxysqualene), (Rowan, 1971). A number of  $\alpha$ -onocerin derivatives have been reported, which belong to the serratene group (Inubushi *et al.*, 1962, Ageta *et al.*, 1962; Tsuda *et al.*, 1970), one of which is serratenediol, isolated from *Lycopodium serratum* by Inubushi *et al.*, (1962). Other molecules such as serratene, onocerane and ambreane, synthesised in the same way have been previously termed onoceroids (Ageta *et al.*, 1982).

Although various research groups have investigated the chemical synthesis of  $\alpha$ -onocerin and its presence in a number of species (table 1.4), the biosynthesis remains to be fully elucidated. The distribution of  $\alpha$ -onocerin and other oxidosqualene cyclase products such as  $\beta$ -amyrin, has been poorly documented with very little information available within the literature. Utilising the metabolomic, genomic and gene expression information available from comprehensively studied plant species may prove to be

important in understanding the complexity of  $\alpha$ -onocerin biosynthesis. This work may help provide background information which could be utilised for future functional expression studies.

### **1.3 TRITERPENES**

#### **1.3.1 Introduction to triterpenes**

The triterpenoid  $\alpha$ -onocerin has been chemically characterised as being part of the ‘isoprenoid pathway’, a term used to describe the biosynthesis of a large group of secondary metabolites. Among the derived groups of compounds are; hemiterpenes, monoterpenes, sequesterterpenes, diterpenes and triterpenes (Bloch, 1992). Within the isoprenoid group, ‘saponins’ and ‘phytosterols’ are derived from the triterpenoid biosynthetic pathway (Phillips *et al.*, 2006). Triterpenes are fused multicyclic structures within which the cyclisation events represent one of the most complex biochemical mechanisms present in nature (Shibuya *et al.*, 2007). Factors affecting the presence of resulting products include; genetic background, tissue type, developmental stage, physiological status and the environment (Dixon *et al.*, 2006).

Phytosterols are derived from the precursors;  $\beta$ -amyrin, cycloartenol, lupeol, stigmasterol, steroidal saponins and triterpene saponins. These precursors can be found across a diverse taxonomic group of plants and animals, in eukaryotes and prokaryotes (Kushiro *et al.*, 1998). Membrane sterols have been implicated in numerous plant functions, regulating membrane permeability and flexibility (Demel and De Kruffyff, 1976) and serving as steroid hormone precursors (Grunwald *et al.* 1975).

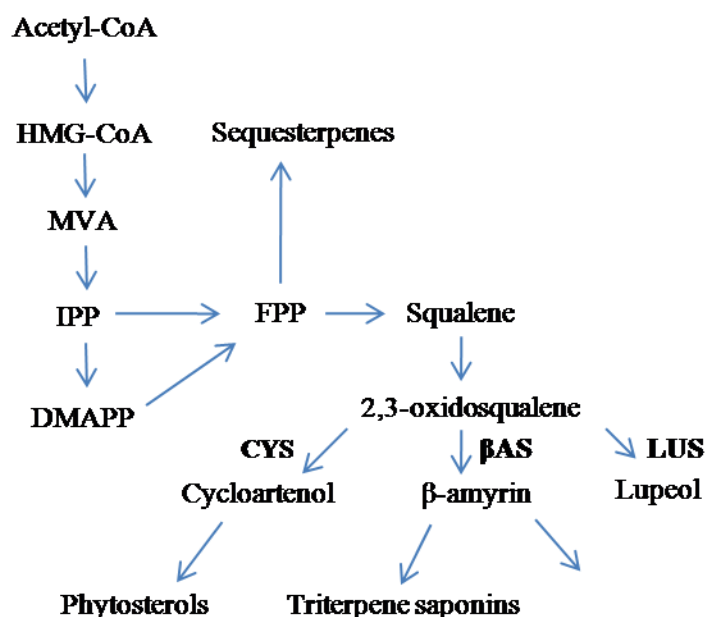
#### **1.3.2 The diversity of plant triterpenes**

A combination of oxidative modifications and glycosylations of carbon skeletons *in planta* has resulted in over 3,000 structurally diverse triterpenes (Harborne, 1999). Triterpenes exhibit a wide range of structural diversity and biological activity which allows them to be exploited in the pharmaceutical industry as antimicrobial agents, making them commercially valuable (Hostettmann and Marston, 1995). Fulton *et al.*, (1994), provided

evidence that triterpene saponins play a role in anti-fungal defence mechanisms. Triterpene saponins are glycosylated triterpenes and have been shown by Papadopoulou *et al.*, (1999) and Haralampidis *et al.*, (2001), to play a role as anti-fungal agents in *Avena strigosa*. The gene involved in encoding the  $\beta$ -amyrin synthase (synthase; an enzyme involved in the coupling of molecules), is a precursor of triterpene saponins, has been shown to have elevated transcription levels between gardenia cultures inoculated with *candida albicans* by Fulton *et al.*, (1994). It has been proposed by Takasaki *et al.*, (1999) triterpenes offer significant potential anti-inflammatory and anti-carcinogen medicinal properties (Balandrin *et al.*, 1985; Takasaki *et al.*, 1999).

$\beta$ -amyrin has been shown to accumulate in plants during the seed germination growth stage, suggesting it has a biological function during early plant development. This has been shown in *Arabidopsis thaliana* (Barton *et al.*, 1975) and was also confirmed by Biasted, (1971) and Hayashi *et al.*, (2001) in other plants, who demonstrated  $\beta$ -amyrin accumulates at higher levels pea and liquorice seedlings, during seed germination.

Triterpenes consist of six isoprene units (30 carbon molecules) and are classified into 6-6-6-5 tetra cycles and 6-6-6-6-5 or 6-6-6-6-6 or 6-6-5-6-6 pentacycles (Xu *et al.*, 2004). Acyclic, monocyclic, bicyclic, tricyclic and hexacyclic triterpenoids are natural products, often playing important roles in defensive mechanisms or as flavanoids (Stafford, 1997). Such a wide range of diversity within a group of secondary metabolites is due to a variation of cyclisation mechanisms, post-cyclic modifications and glycosylations (Phillips *et al.*, 2006). In contrast to the large number of products, triterpenes are derived from a small number of precursors, with the initial starting substrate for all triterpenes being squalene.



**Figure 1.3** Triterpene biosynthetic pathway: Acetyl Co-enzyme A, 3-hydroxy-3-methylglutaryl-coenzyme A, mevalonic acid (MVA), isopentenyl diphosphate (IPP), dimethylallyl diphosphate (DMAPP), farnesyl diphosphate (FPP), cycloartenol synthase (CYS),  $\beta$ -amyrin synthase ( $\beta$ AS), lupeol synthase (LUS). (Suzuki *et al.*, 2002).

The mechanism of triterpene cyclisation was first described by Ruzicka in the 1950s, as ‘an unusual and remarkably flexible reaction mechanism’ (Ruzicka *et al.*, 1953). A set of rules were deduced, which could explain the biosynthesis of each triterpene skeleton, known as the biogenetic isoprene rules. Isoprenes (5 carbon cyclic rings) are the basic units that these compounds are composed from. By utilising the known structure and the basic set of principles outlined in the biogenic isoprene rule, an accurate model of triterpene cyclisation has been developed (Ruzicka, 1959).

### 1.3.3 Triterpene biochemical pathway

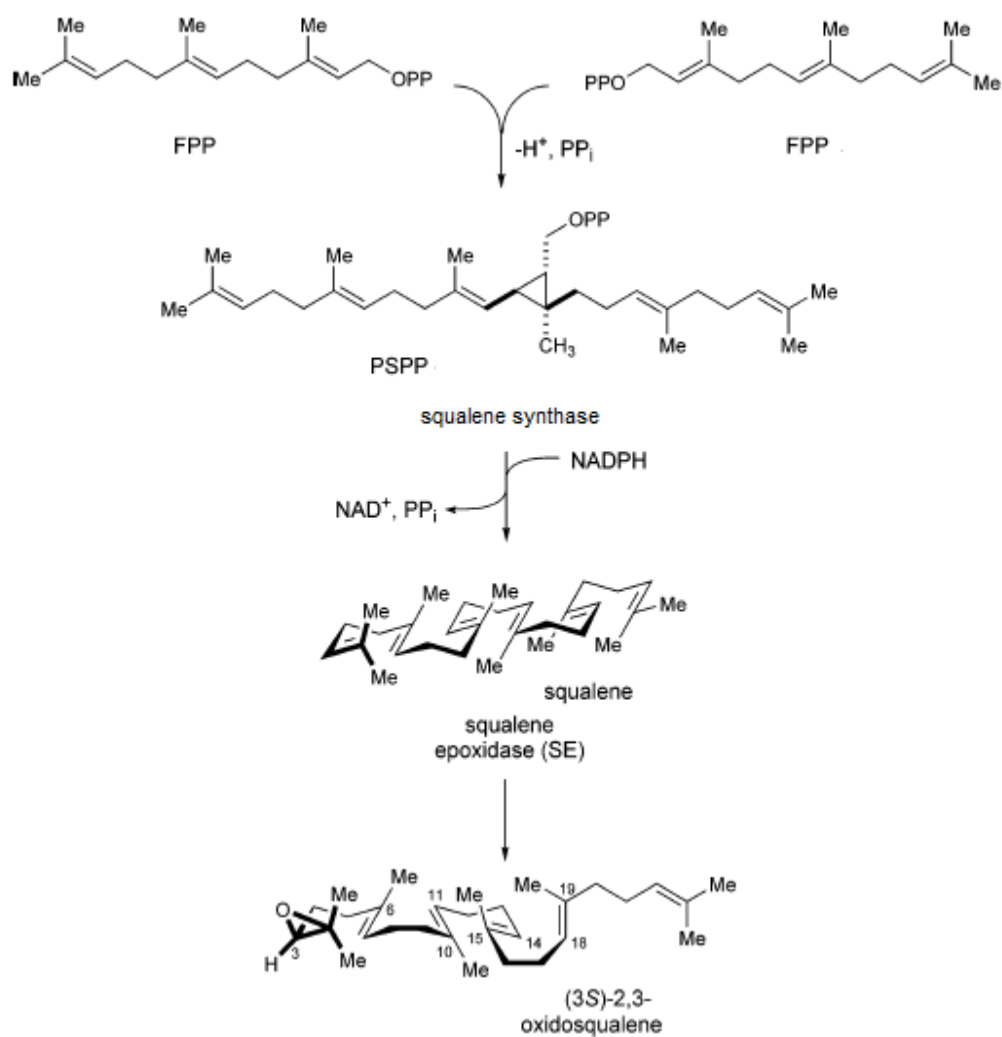
The first step towards triterpene biosynthesis as illustrated in figure 1.3 involves the homo-coupling of two farnesyl diphosphate molecules to produce a pre-squalene pyrophosphate (PSPP) through the action of squalene cyclase. Squalene is a symmetrical molecule with a C<sub>30</sub> backbone and is considered as the primary precursor for all triterpenes (Xu *et al.*,

2004). The next stage is an enantio-selective epoxidation event of squalene by squalene epoxidase to produce 2,3-oxidosqualene also shown in figure 1.4 (Yoder and Johnson, 2005).

Triterpenes are assigned as either squalene cyclase products or 2,3-oxidosqualene products. Some 2,3-oxidosqualene cyclase products can also be derived from squalene via a different cyclisation mechanism, followed by an oxidation step by cytochrome P450 oxidation enzymes at the C<sub>3</sub> position (Xu *et al.*, 2004). Alternatively squalene products can originate from 2,3-oxidosqualene with the 3-hydroxyl group becoming reduced after cyclisation by cytochrome P450 hydroxylation enzyme (Xiong *et al.*, 2006). It is then possible for these single substrates to be converted into multiple products (Husselstein-Muller *et al.*, 2001). It has also been documented that multiple products can be partially cyclised and reaccepted into the active site to produce a diverse range of triterpene precursors (Husselstein-Muller *et al.*, 2001).

Often due to the lack of enzyme substrate specificity, bi-products are produced within a single cyclisation event (Haralampadis *et al.*, 2002). Due to such diversity in enzyme functionality within the triterpene pathway, it is difficult to predict cyclisation mechanisms without full biochemical and genetic characterisation of gene function. To date, for this reason  $\alpha$ -onocerin has only been characterised as a product of the triterpene pathway and it is not known whether squalene is a precursor of  $\alpha$ -onocerin, or whether  $\alpha$ -onocerin accumulates as a consequence of further cyclisation of squalene.





**Figure 1.4** Schematic diagram showing the stages involved in the biosynthesis of 2,3-oxidosqualene. Two farnesyl diphosphates (FPP), combine by losing two phosphates and one hydrogen to produce a pre-squalene pyrophosphate (PSPP). A pre-squalene pyrophosphate (PSPP) is formed prior to losing two phosphates by the action of squalene synthase to produce squalene. Squalene epoxidase adds an epoxide group (OH) on the 2,3 carbon terminal producing squalene epoxide (adapted from Yoder and Johnson, 2005).

#### 1.4 OCCURRENCE AND SUGGESTED FUNCTION OF $\alpha$ -ONOCERIN

Although there have been numerous investigations into the stereochemistry (Barton and Overton 1955; Rowan 1971; Pauli and Fröhlich 2000) and synthesis (Mi *et al.*, 2002) of  $\alpha$ -onocerin there is little evidence regarding the ecological function.  $\alpha$ -Onocerin and onoceroids have previously been isolated from a very diverse range of plants, including angiosperms, gymnosperms, ferns and club mosses which may indicate a level of adaptive convergence (table 1.4).

Recent paleontological studies have contributed towards our understanding of the potential functions of  $\alpha$ -onocerin. Onocerane I has been detected in sediments deposited during the two driest periods in northern Brazil (Last Glacial maximum and Younger Dryas. Jacob *et al.*, (2004) suggested that onocerane I might be an indicator that plants had adapted to arid and semi-arid environments. Palaeoclimatic and palynological results suggest water scarcity and an absence of pollen from water associated plants such as ferns and club mosses in these areas (Jacob *et al.*, 2004).

In North East Brazil, high levels of onocerane I were recorded 13,000 years before present. Paleontological data confirmed ferns and club mosses were not present at that time. The fact that Restharrow has not been characterised as being present in NE Brazil either presently or historically, suggests an alternative source of  $\alpha$ -onocerin. During this period a regression of moist forest occurred towards open vegetation. The presence of *Cecropia* species, (a pioneering species colonising following the decline in humid forests), indicated abrupt climate change (Jacob *et al.*, 2004).

**Table 1.4** The table describes molecules related to  $\alpha$ -onocerin, onoceroid derivatives and the plant species in which they are present. References are listed and adapted from Jacob *et al.*, 2004.

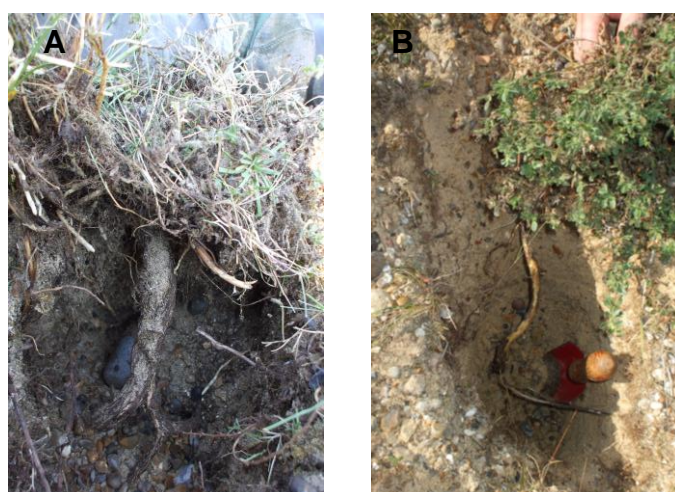
Secondary metabolite	Plant species	Reference
Colysanoxide	<i>Colysis</i>	Ageta <i>et al.</i> , 1982
Onoceradienes	<i>Lemmaphyllum microphyllum</i> (Polypodiophyta)	Ageta <i>et al.</i> , 1982
Ononceranoxide		Matsuda <i>et al.</i> , 1989
Onocer-7-ene-3,21-diol	<i>Cissus quadrangularis</i> (Rhamnales, Vitaceae)	Bhutani <i>et al.</i> , 1989
Lansiosides	<i>Lansium domesticum</i> (Sapindales, Meliaceae)	Nishizawa <i>et al.</i> , 1983
Lansic acid		Habaguchi <i>et al.</i> , 1968
Onocerane II and III Onocerane I	<i>Pseudofagus</i> sp. (Fagaceae)	Gianasi Niklas 1985 Jacob <i>et al.</i> , 2004
$\alpha$ -onocerin	<i>Ononis</i> sp.(Fabaceae)  <i>Lycopodium clavatum</i> , <i>L. obscurum</i> , <i>L. deuterodensum</i> (Lycopodiophyta). <i>L. sitchense</i> <i>L. inundatum.</i> , <i>L. obscurum</i> , <i>L. asuarinoides</i> , <i>Palhinhaea cernua</i> <i>Glycyrrhiza glabra</i> <i>Melicope perspicuinervia</i>	Barton and Overton 1955, Rowan and Dean Tsuda <i>et al.</i> , 1984 Inubushi <i>et al.</i> , 1971 Cai <i>et al.</i> , 1992 Bisset <i>et al.</i> , 2001 Murphy <i>et al.</i> , 1974

$\alpha$ -onocerin and other related onoceroids may serve a very different role in plants associated with humid and saturated environments, such as club mosses and ferns. In contrast, Batanouny (2001) argues  $\alpha$ -onocerin may

play an important role in protecting higher plants in water deficient environments.

Saline habitats are principally characterized by high sodium chloride levels and high osmotic potential shadowed by high concentrates of sulphates, carbonates, poor soil structure and generally high soil pH (Batanouny, 2001). There are two types of salt stress which affect the growth response of plants. Osmotic stress, is where the growth medium has a lower water potential than the plant tissue and ionic stress is where the concentration of sodium chloride ions present within the surrounding environment is higher than within the plant. Saline conditions often inhibit the growth of crop plants, initially affecting root systems which are in direct contact with the soil. Adaptations could occur in the plant as a result of these environmental pressures.

The composition of the root membrane can affect the flow of sodium and chloride ions between the root cells and the soil. It is possible that in the case of *Ononis* species,  $\alpha$ -onocerin inhibits the flow of sodium chloride ions and allows the root to draw fresh water from below the saline level.



**Figure 1.5** Photographs of *O. repens subsp. maritima* in their natural environment (Hayes, 2008). Image A) shows the robust woody tap root, B) deep tap root growing on sandy shingle situated in a shingle habitat (Dunwich).

Rowan and Dean (1972b) quantified  $\alpha$ -onocerin in 12 *Ononis* species. The authors found  $\alpha$ -onocerin to be present in greater quantities in the root system than in the aerial parts of the plant within most Restharrow species (table 1.5).

According to Batanouny, (2001) deserts are areas with less than 250 mm of annual rainfall, consisting of sparse vegetation made up of plants which have evolved particular characteristics to enable them to endure saline soils and drought. These adaptations can be in the form of physiological function such as modified metabolic pathways, altered life cycles, or adaptive morphological characteristics (Aronson *et al.*, 1992).

*Ononis vaginalis*, a woody perennial, is an early coloniser of sand dunes in northern Africa, persisting in moderately saline soil ( $0.25\text{-}0.83\text{ ms/cm}^{-1}$ ), (Migahid, 2003). *Ononis vaginalis* has developed several characteristics to tolerate desert conditions, which are; a large tap root, crassulacean acid metabolism (CAM), small leaves, and leaf hairs to reduce transpiration (Migahid, 2003). *Ononis repens subsp. maritima* is subject to similar conditions in coastal regions of the UK, although temperature ranges are not as extreme and water availability *per se* is not limiting.

Mahanesh and El-Oqlah (1999) suggest  $\alpha$ -onocerin as potentially playing a functional role in plant defence. Results from antimicrobial activity experiments showed weak antifungal activity and weak antimicrobial activity against *Escherichia coli*, *Staphylococcus typhimurium*, *Staphylococcus aureus*, *Bacillus cereus*, *Aspergillus flavus*, and *Candida albicans*, with limited inhibition of bacterial growth (Mahanesh and El-Oqlah, 1999).

**Table 1.5** The table displays the total  $\alpha$ -onocerin content in milligrams per gram (mg/g) dry weight, isolated from whole tissue, root and aerial hexane extracts of 12 *Ononis* species. The analytes were identified and quantified using Gas chromatography/mass spectrometry GC/MS Trace indicates insufficient quantity of  $\alpha$ -onocerin to measure (Rowan, 1971).

Restharrow ( <i>Ononis</i> ) species	Plant tissue type	Quantity (mg/g dry weight)
<i>O. cristata</i>	Root	5.70
	Aerial	0.20
<i>O. viscosa</i>	Root	0
	Aerial	0.031
<i>O. pubescens</i>	Root	0.50
	Aerial	0.036
<i>O. reclinata</i>	Root	0
	Aerial	0
<i>O. pusilla</i>	Root	0
	Aerial	0
<i>O. minutissima</i>	Root	0.12
	Aerial	0
<i>O. arvensis</i>	Root	5.39
	Aerial	0.25
<i>O. spinosa</i>	Root	4.12
	Aerial	0.54
<i>O. repens</i>	Root	2.28
	Aerial	0.025
<i>O. alopecuroides</i>	Root	2.79
	Aerial	Trace
<i>O. mitissima</i>	Root	1.22
	Aerial	0.019
<i>O. subspicata</i>	Root	0.35
	Aerial	0



**Figure 1.6** Rhizobial nodules on the roots of *Ononis spinosa* (Peter G D Dean, 2008).

Restharrow may provide a suitable habitat for bacterial growth within the roots. Rhizobia nodules can form on the tap and lateral roots of *O. spinosa* and *O. repens*. The Rhizomes may provide a habitat for *Agrobacterium* to occupy, and potentially form symbiotic relationships with nitrogen fixing bacteria (figure 1.6).

There is no evidence in the literature to suggest Restharrows are grazed upon by mammals, although they play a role providing food for primary consumers such as Weevil larvae *Sitona ononidis*, *Hypera nigrirostris* (Coleoptera), (Bullock, 1992) and leaf miner larvae, *Liriomyza cicerina* (Diptera), (Spencer, 1972). The seeds also provide a useful refuge and food source for *Apion ononis* (Coleoptera), a type of seed weevil (Morris, 1990). There is currently no direct evidence to suggest onocerin has anti-herbivory activity.

Shallari *et al.*, (1998), described *Ononis spinosa* as being a heavy metal tolerator but not a hyperaccumulator. One possible explanation suggested by Shallari *et al.*, (1998) is that  $\alpha$ -onocerin may act as a wax, inhibiting the flow of toxic ions into the root system.

### 1.5 MEDICINAL USES OF $\alpha$ -ONOCERIN

There are reports of plant extracts from *Ononis* species being used for medicinal purposes and as having diuretic, expectorant and antiseptic potential (Mahasneh and El-Olqah, 1999; Jaric *et al.*, 2007). It is unclear whether,  $\alpha$ -onocerin is the active compound (Mahasneh and El-Oqlah, 1999). Orhan *et al.* (2002) have investigated the possibility of using  $\alpha$ -onocerin as an Acetylcholine esterase inhibitor. The inhibition effect of  $\alpha$ -onocerin isolated from *Lycopodium clavatum*, of acetylcholine esterase was comparable to donepezil (an existing treatment against Alzheimer's). At 3 mg/ml,  $\alpha$ -onocerin had an inhibition rate of 33.39% compared to 20.05% for donepezil. This suggests that  $\alpha$ -onocerin may be an active inhibitor of acetylcholine esterase, and could potentially be used as a treatment of Alzheimer's disease. In contrast a more recent report by Rollinger's group (2005) discovered that  $\alpha$ -onocerin isolated from *O. spinosa* showed no inhibitory effect using Ellman's reagent in a microplate assay.

Recent studies to investigate the potential medicinal uses of  $\alpha$ -onocerin include that of Yolmaz *et al.*, (2006). The analgesic properties of *O. spinosa* extracts were determined based on dose responsive pain thresholds of mice, measured using a tail flick response test. The analgesic activity of the extracts was equal to that of aspirin when administered at a 100 mg dose. Contrary to the hearsay evidence in books and on web sites, Yolmaz *et al.*, (2006) showed there was no hepatoprotective influence of *O. spinosa* extracts, when liver tissue was induced with carbon tetrachloride, aspartate amino transferase, alanine amino transferase and bilirubin. During an attempt to identify the active constituents from *Lycopodium japonicum* one of the most common herbal medicines in China for use against arthritic pain, dysmenorrhoea and contusion,  $\alpha$ -onocerin was isolated. It was unclear whether  $\alpha$ -onocerin was the primary active compound (Inubushi *et al.*, 1997).



## 1.6 BIODIVERSITY OF 2,3-OXIDOSQUALENE CYCLASES IN PLANTS

The biological functions of triterpenes and their cyclic enzymes are generally poorly understood. Genes have been characterised, using chemical, molecular and genome mining techniques (Suzuki *et al.*, 2002). The sequencing of *Arabidopsis thaliana* and *Oryza sativa* has contributed towards our understanding of triterpene biosynthesis within plants. Thirteen oxidosqualene cyclases in *A. thaliana* and nine oxidosqualene cyclases in *O. sativa* have been characterised to date. Functional characterisation studies of triterpene synthases such as cycloartenol and lanosterol synthase, have been carried out across eight different plant species one of which is *Pisum sativum*. Table 1.6, 1.7, 1.8 and 1.9 show the 2,3-oxidosqualene cyclases which have been characterised within plants to date.

Previous functional studies of triterpene synthases have revealed two classes of 2,3-oxidosqualene cyclases in plants. The first class are mono-functional synthases, classified as yielding single triterpenoid products. Some of these have been characterised by over-expression of triterpene synthases in *saccharomyces cerevisiae*. Examples of characterised single product synthases are outlined in table 1.6, 1.7 and 1.8. Single substrate synthases represent a branch in the triterpene pathway and are either  $\beta$ -amyrin synthases (Hayashi *et al.*, 2004; Iturbe-Ormaetxe *et al.*, 2003; Kajikawa *et al.*, 2005; Kushiro *et al.*, 1998; Moritita *et al.*, 2003; Sawai *et al.*, 2006; Zhang *et al.*, 2002), or Lupeol synthases (Guhling *et al.*, 2006; Hayashi *et al.*, 2004; Shibuya *et al.*, 1999; Zhang *et al.*, 2003).

**Table 1.6** Characterised 2,3-oxidosqualene cyclase: cycloartenol synthase within plant species. The gene name and the primary plant metabolomic product are listed. Analytical methods used for characterisation studies; Gas chromatography (GC), Mass Spectrometry (MS), High Pressure Liquid Chromatography (HPLC), Liquid Chromatography (LC) and Nuclear Magnetic Resonance (NMR). Adapted from Herrera (1999) and Phillips *et al.*, (2006).

Plant species	Gene name	Secondary metabolite product	Identification method
<i>Abies magnifica</i>	CAS1, AG44096	Cycloartenol	NMR,GC
<i>A. thaliana</i>	AT2G07050/CAS1 P38605	Cycloartenol	NMR,IR,MS
<i>B. platyphylla</i>	BPX1 BAB83085	Cycloartenol	LC-MS
<i>B. platyphylla</i>	BPX2, BAB83086	Cycloartenol	LC-MS
<i>C. speciosus</i>	CSOSC1, BAB83253	Cycloartenol	LC-MS
<i>C. pepo</i>	CPX	Cycloartenol	LC-MS
<i>G. glabra</i>	GgCAS1, BAA76902	Cycloartenol	-
<i>L. cylindrical</i>	LcCAS1, BAA85266	Cycloartenol	-
<i>P. gensing</i>	PNX, BAA33460	Cycloartenol	HPLC
<i>P. sativum</i>	CASPEA, BAA23533	Cycloartenol	-

The second group of triterpene synthases are multifunctional enzymes, cyclising triterpenoid products from 2,3-oxidosqualene (known as oxidosqualene cyclases), (table 1.9). Examples of multifunctional enzymes include; Lupeol synthase, (AT1G78970, LUP1), cyclising six triterpenoid products and  $\beta$ -amyrin synthase (At1g78960), which is known to cyclise nine triterpenoid products.

**Table 1.7** Characterised 2,3-oxidosqualene cyclase:  $\beta$ -amyrin synthase within plant species. The gene name and the primary plant metabolomic product are listed. Analytical methods used for characterisation studies; Gas chromatography (GC), Mass Spectrometry (MS), High Pressure Liquid Chromatography (HPLC), Liquid Chromatography (LC) and Nuclear Magnetic Resonance (NMR). Adapted from Herrera (1999) and Phillips *et al.*, (2006).

<i>Plant species</i>	Gene name	Secondary metabolite product	Identification method
<i>A. strigosa</i>	ASBAS1, CAC84558	$\beta$ -amyrin	HPLC
<i>B. platyphylla</i>	BPY, BAB830088	$\beta$ -amyrin	LC-MS
<i>E. tirucalli</i>	ETAS, BAE43642	$\beta$ -amyrin	GC-MS
<i>G. glabra</i>	GgBAS1, BAA89815	$\beta$ -amyrin	LC-MS
<i>M. truncatula</i>	MtAMY1/BAS1, CAD23247	$\beta$ -amyrin	GC-MS, NMR, HPLC
<i>P. ginseng</i>	PNY1, BAA33461	$\beta$ -amyrin	HPLC, MS, NMR
<i>P. ginseng</i>	PNY2, BAA33722	$\beta$ -amyrin	-
<i>P. sativum</i>	PSY, BAA97558	$\beta$ -amyrin	LC-MS, GC-MS, NMR

Seven triterpenoids have been characterised as products from *Pisum sativum* (pea) mixed amyrin synthase (PSM-BAA97559), (Moritia *et al.*, 2000). To date there are more than 100 known functional triterpene synthases from within plants. (Segura *et al.*, 2000).

Mutagenesis studies have shown that single base pair changes can alter enzyme specificities, leading to the production of alternative oxidosqualene cyclase products. Such mutations are responsible for producing the estimated one hundred 2,3-oxidosqualene-derived products which have so far been elucidated (Xu *et al.*, 2004). This catalytic flexibility has presumably contributed to the diverse range of triterpenes observed in nature. It is therefore difficult to predict product output simply by analyzing

nucleic acid sequences and phylogenetic relationships. Within *A. thaliana*, lupeol synthase (At1g78970), converts 2,3-oxidosqualene into lupeol,  $\beta$ -amyrin, germanicol, taraxasterol,  $\psi$ -taraxasterol and 3,20-dihydroxylupane (Herrera *et al.*, 1998; Segura *et al.*, 2000).

A related *A. thaliana* multi-functional triterpene synthase (At1g78960), produces lupeol, taraxasterol,  $\beta$ -amyrin,  $\psi$ -taraxasterol, baurenol,  $\alpha$ -amyrin and multiflorenol (Kushiro *et al.*, 2000; Husselstein-Muller *et al.*, 2001). The two enzymes cyclise initially to the lupyl cation and then allow further rearrangements to germanicyl and taraxastyl cations, quenched by deprotonations.

In an attempt to understand triterpene synthesis and function at least nine different triterpene synthase genes have so far been cloned. These include  $\beta$ -amyrin synthases from *Panax ginseng* (PNY1, BAA33461, PNY2, BAA33722), (Kushiro, 1998), *Arabidopsis thaliana* (At1g78950 and At1g78960), (Herrera *et al.*, 1998), *Taraxicum officinale* (Shibuya *et al.*, 1999), lupeol synthases from *Olea europea* (Shibuya *et al.*, 1999), *Pisum sativum* (Morita *et al.*, 2000), *Glycyrrhiza glabra* (Hayashi *et al.*, 2000) *Medicago truncatula* (Suzuki *et al.*, 2002), *Lotus japonicus* (Iturbe-Ormaetxe *et al.*, 2003), and *Euphorbia tirucalli* (Kajikawa, 2005).

Originally characterised from *A. thaliana* by cloning the cDNA into cycloartenol synthase (CAS) yeast mutants (Corey *et al.*, 1993) 2,3-oxidosqualene is converted to cycloartenol by cycloartenol synthase (CAS) through the protosteryl cation intermediate. From this approach homology based isolation of CAS was continued in a number of eudicots, gymnosperms and monocots.

**Table 1.8** Characterised 2,3-oxidosqualene cyclase: lupeol synthase within plant species. The gene name and the primary plant metabolomic product are listed. Analytical methods used for characterisation studies; Gas chromatography (GC), Mass Spectrometry (MS), High Pressure Liquid Chromatography (HPLC), Liquid Chromatography (LC) and Nuclear Magnetic Resonance (NMR). Adapted from Herrera (1999) and Phillips *et al.*, (2006).

Plant species	Gene name	Secondary metabolite product	Identification method
<i>B. platyphylla</i>	BPW, BAB83087	Lupeol	LC-MS
<i>G. glabra</i>	GgLUS1, BAD08587	Lupeol	LC-MS
<i>O. europaea</i>	OEW, BAA86930	Lupeol	LC-MS
<i>T. officinale</i>	TRW, BAA86932	Lupeol	NMR

The similarity with CAS genes from amoebae and bacteria suggests CAS predates the emergence of plants. Lanosterol is the initial acyclic precursor for sterols in animals and fungi, however there have been instances where lanosterol biosynthesis has been demonstrated in plants. One example is in the latex of *Euphorbia* species, which contains lanosterol. There are also homologues in *A. thaliana* which are similar to CAS1 (65% identity at the amino acid level), which have recently been characterized as evolving independently between plants and animals (At3g45130), (Kolesnikova *et al.*, 2006).

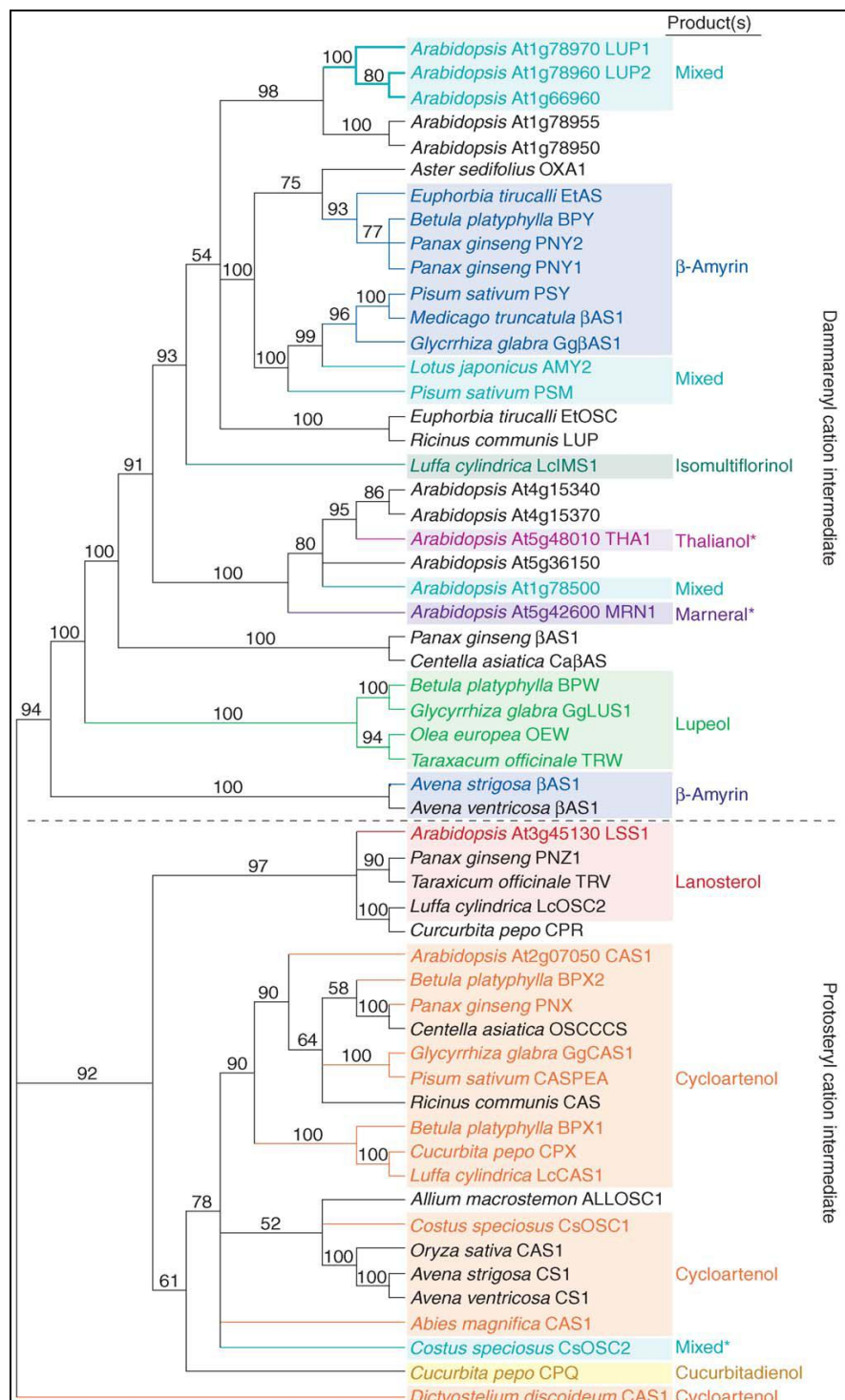
According to Kolesnikova *et al.* (2006), CAS1 can be mutated to cyclise the same products as lanosterol synthase (LAS) with only two amino acid substitutions: Histidine (His) at position 477 to Asparagine (Asn) and Isoleucine (Ile) at position 481 to Valine (Val). This suggests only small evolutionary changes would be necessary to convert CAS to lanosterol synthase (LAS) if natural selection favored lanosterol biosynthesis (Kolesnikova *et al.*, 2006).

**Table 1.9** Characterised 2,3-oxidosqualene cyclase: mixed oxidosqualene cyclases within plant species. The gene name and the primary plant metabolomic product are listed. Analytical methods used for characterisation studies; Gas chromatography (GC), Mass Spectrometry (MS), High Pressure Liquid Chromatography (HPLC), Liquid Chromatography (LC) and Nuclear Magnetic Resonance (NMR). Adapted from Herrera (1999) and Phillips *et al.*, (2006).

Plant species	Gene name	Secondary metabolite product	Identification method
<i>A. thaliana</i>	AT1G78970, LUP1	Lupeol, 3 $\beta$ ,20-dihydroxylupane, $\beta$ -amyrin, germanicol, taraxasterol and $\psi$ -taraxasterol	GC-MS, NMR
<i>A. thaliana</i>	AT1G78960, LUP2	$\beta$ -amyrin, taraxasterol, tirucalla-7,21-dien-3 $\beta$ -ol, lupeol, bauerenol, butyrospermol, multiflorenol, alpha-amyrin and $\psi$ -taraxasterol	HPLC, GC-MS, NMR, LC-MS
<i>A. thaliana</i>	AT1G66960	Tirucalla-7,21-dien-3 $\beta$ -ol	LC-MS
<i>A. thaliana</i>	AT1G78500/PEN6	Bauerenol, lupeol, amylin	LC-MS
<i>C. speciosus</i>	CsOSC2, BAB83254	Lupeol, germanicol, $\beta$ -amylin, Lupeol,	LC-MS
<i>L. japonica</i>	AMY2, AAO33580	Lupeol, $\beta$ -amylin	HPLC
<i>P. sativum</i>	PSM, BAA97559	$\psi$ -taraxasterol,, lupeol, butyrospermol, germanicol, taraxasterol, $\beta$ -amylin, $\alpha$ -amylin	LC-MS, GC-MS, NMR

Cucurbitadienol synthase has arisen from duplication which has 65% and 71% amino acid identity to CAS and LAS, respectively. Cucurbitadienol synthase produces cucurbitacin, a bitter compound, which acts as an anti-feedant (Phillips *et al.*, 2006). In order for cucurbitadienol to be biosynthesised the protosteryl cation is required rather than the dammarenyl cation.

Products from the dammarenyl cation are often produced from OSC's which produce single products. Lupeol synthases, cyclise 2,3-oxidosqualene to the dammarenyl cation which promotes cyclic ring expansion and annulation to the lupyl cation and terminates by abstracting the C-29 proton to form lupeol.  $\beta$ -amyrin as mentioned previously, is also formed from the lupyl cation, but allows further ring expansion and rearrangement before deprotonation to  $\beta$ -amyrin. Harampidis *et al.*, (2001), suggested that  $\beta$ -amyrin synthases are distinctly less similar (50% identity) to each other than CAS enzymes (70-79% identity), which highlights the possibility of independent origins of  $\beta$ -amyrin synthases in eudicots and monocots (Harampidis *et al.*, 2001; Qi *et al.*, 2004).



**Figure 1.7** An unrooted phylogram showing the relationship of triterpene synthases (cycloartenol synthase, β-amyirin synthase, lupeol synthase, and multifunctional synthases) between species. The phylogram identifies two distinct groups of plants where triterpene products are derived from protosteryl and dammarane cations (Phillips *et al.*, 2006).

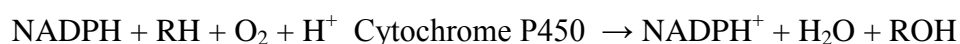


The diverse metabolites derived from  $\beta$ -amyrin and lupeol precursors are involved in various plant processes in different ways.  $\beta$ -amyrin is a precursor of triterpene glycosides, such as antifungal avenacin found in *A. strigosa* root.  $\beta$ -amyrin metabolites accumulate in specific tissues, as specified in *Pisum*, *Medicago*, *Lotus*, *Avena* and *Glycyrrhiza* (Harampidis *et al.*, 2001).

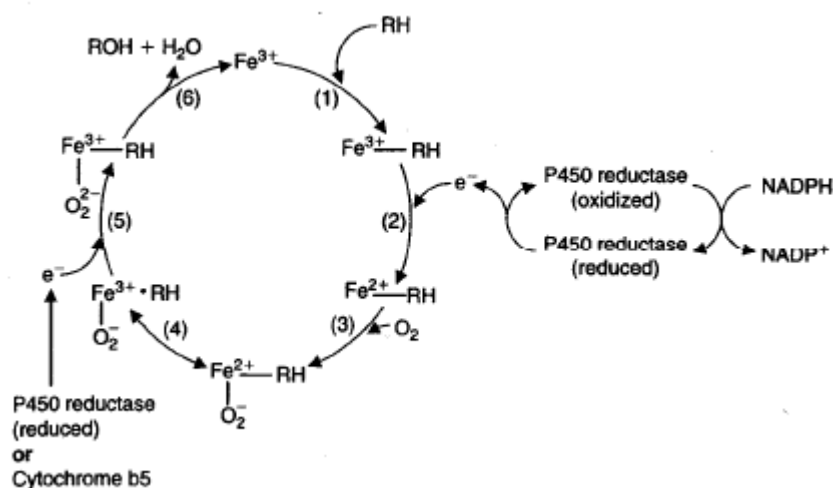
Interestingly a number of plant triterpene synthases use the dammarenyl cation, to produce multiple products in legumes. These multifunctional triterpene synthases share a common ancestor with uni-functional  $\beta$ -amyrin synthases (Harampidis *et al.*, 2001), however other *A. thaliana* mixed amyrin synthases (multifunctional OSC enzymes) have arisen repeatedly during angiosperm evolution. An example is *Costus speciosus*, which biosynthesises lupeol, germanicol and  $\beta$ -amyrin from the dammarenyl cation. The closest relatives are cycloartenol synthases (65-70% identity) which use the protosteryl cation. This relationship implies a recent evolutionary transition that involved an aromatic catalytic change from the prosteryl cation to the dammarenyl cation (Xiong *et al.*, 2006).

## 1.7 CYTOCHROME P450 HEMOPROTEINS

Cytochrome P450 enzymes are a super family of hemoproteins present in most living organisms, both plants and animals (Nelson, 1981) which catalyse the oxidation of metabolic substrates. First discovered by Klingenberg and Garfinkle (1958) as a new pigment in rat liver, they were first characterised by Omura and Sato in 1962. To date there are 400 known members of the cytochrome P450 family (Nelson, 1981). The name P450 is derived from the fact that, at a reduced state P450 enzymes bind carbon monoxide, giving an absorption maximum of 450nm. (Garfinkle 1958, Jefcoate, 1978; Guengerich, 1991). The mixed function reaction consists of cytochrome P450, NADPH-450 (Nicotinamide adenine dinucleotide phosphate), cytochrome reductase and can be summarised as outlined on page 30.



Cytochrome P450's usually occur in plants as membrane associated forms in the endoplasmic reticulum and are approximately five hundred amino acids in length at 60KDa (Hocch and Montellano, 2001). Typically the main function of cytochrome P450 is the mono-oxygenation of carbon (R) substrates, within the catalytic cycle with the best known model to date, proposed by (Guengerich 1991; Guengerich and Johnson 1997; De Voss *et al.*, 1997). Cytochrome P450 participates in a wide variety of reactions including epoxidations, dealkylations, deaminations and dehalogenations. P450 enzymes are substrate dependant and catalyse many stages in sterol and fatty acid biosynthesis. Oxidation and hydroxylation of substrates can alter the polarity of metabolic precursors, allowing intermediate precursors to become more susceptible to absorption by active sites of other enzymes (Lewis *et al.*, 1997; Nelson, 1981).



**Figure 1.8** Action of cytochrome P450 catalytic cycle. Six stages in the catalytic cycle;  $\text{Fe}^{2+}$ =oxidized cytochrome P450,  $\text{Fe}^{3+}$ = reduced cytochrome P450, b5 = cytochrome b5, RH = substrate, ROH= product and  $\text{e}^-$ = electron (adapted from Meunier *et al.*, 2004).

P450 hydroxylation enzymes are named according to the carbon on which the active site alters. In *Arabidopsis dwf4*, a brassinosteroid (BR)-deficient mutant, was used to characterise the function of a steroid C22-hydroxylase

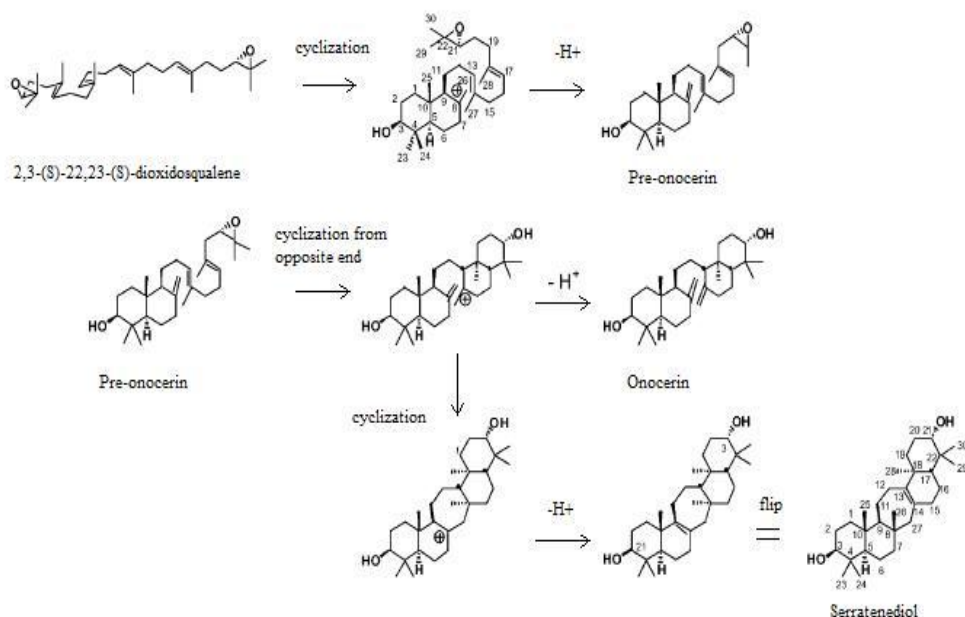
(CYP90B1). The CYP90B1 converted campestanol to 6-deoxocathasterone, confirming that CYP90B1 is a steroid C-22 hydroxylase. The substrate specificity of CYP90B1 indicated that sterols with a double bond at positions C-5 and C-6 are preferred substrates compared with stanols, which have no double bond at the position within the brassinosteroids pathway (Fujita *et al.*, 2006). NADPH-cytochrome P450 Reductase and supernatant protein cofactors have been shown to function with 2,3-oxidosqualene cyclase (Abe and Prestwich, 1999) and are thought to have rate limiting effects on the accumulation of 2,3-oxidosqualene (Shibata *et al.*, 2001; Meunier *et al.*, 2004).

## **1.8 PROPOSED MECHANISTIC PATHWAYS FOR $\alpha$ -ONOCERIN BIOSYNTHESIS**

Although elucidation of chemical synthesis is important in terms of understanding stereochemistry and predicting cyclic events, it is also important to understand how molecules are formed within biological systems. Several mechanisms have been proposed for the biosynthesis of  $\alpha$ -onocerin. The first committed step towards  $\alpha$ -onocerin biosynthesis is initiated by a squalene synthase combining two geranyl pyrophosphate molecules to form squalene. Squalene epoxidase adds an epoxide (OH) group. From this point it is unclear how  $\alpha$ -onocerin is synthesised in nature. In other organisms triterpene precursors such as epoxydammerenes have been proposed to arise from tetracyclisation of 2,3-oxidosqualene, followed by P450 oxidation towards E ring formation. Shaun *et al.* (2005) investigated the theory that oxidation of 2,3-oxidosqualene by squalene epoxidase, before cyclisation may occur offering an alternative evolutionarily accessible mechanism. As discussed previously, the group proposed; 2,3;22,23-dioxidosqualene may be converted by a triterpene synthase directly in a single cyclisation step. For example during cholesterol biosynthesis a 2,3;22,23-dioxidosqualene (ubiquitous) bi-product of 2,3-oxidosqualene, has been reported as a rate limiting step towards cholesterol biogenesis.

Work carried out by Rowan (1971), proposed that  $\alpha$ -onocerin could be cyclised directly from a 2,3;22,23-dioxidosqualene intermediate. Enzyme assays containing squalene synthase and oxidosqualene cyclase accumulated  $\alpha$ -onocerin when spiked with squalene. It is a plausible alternative for cyclisation of four ring cyclic structures. Incomplete cyclisation from the three carbon terminal can lead to intermediates as discussed previously; these have subsequently been shown to undergo a further cyclisation event from the C21/22 distal epoxide. There is no direct evidence to suggest whether this is as a result of re-entry into the same enzyme/substrate complex, a dual cyclisation event or caused by an additional oxidosqualene cyclase.

According to the literature  $\alpha$ -onocerin (Barton and Overton, 1955) is a product of 2,3;22,23-dioxidosqualene that arises by separate bicyclisation from each terminus. The biosynthesis of this bis-bicyclic triterpene diol from 2,3;22,23-dioxidosqualene has been established in a cell-free system from *O. spinosa* (Rowan *et al.*, 1971). Protonating one epoxide would initiate bicyclisation and cause proton abstraction from C26 yielding a pre-onocerin (Rowan and Dean, 1972). Protonating the distal epoxide would initiate a second bicyclisation, and deprotonation generating  $\alpha$ -onocerin. Potential genes which may be responsible for encoding oxidosqualene cyclases which subsequently cyclise 2,3;22,23-dioxidosqualene into  $\alpha$ -onocerin include; lupeol synthase,  $\beta$ -amyrin synthase or cycloartenol synthase. The most likely explanation is that one of these enzymes would play a multifunctional role. To date a pre-onocerin substrate has failed to be detected, suggesting a single cyclisation event acting on 2,3;22,23-dioxidosqualene substrate may be the most likely pathway, although the pre-onocerin could potentially be unstable or be rapidly consumed, preventing detection.



**Figure 1.9** Proposed mechanistic pathway of triterpene biosynthesis. 2,3;22,23-dioxidosqualene cyclises initially to form a pre-onocerin intermediate. Further modification produces a group of compounds found in pine species, serratine diols. (Shuan *et al.* 2005).

## 1.9 POTENTIAL CYCLIC ROUTES TOWARDS $\alpha$ -ONOCERIN BIOSYNTHESIS

The proposed mechanisms are based on cyclisation events of either 2,3-oxidosqualene or 23,22:2,3-dioxidosqualene (shown in figure 1.9) as a direct precursor of  $\alpha$ -onocerin. Given the complex nature and diversity of biological, genetic and biochemical questions, there are several approaches, which could be taken to investigate  $\alpha$ -onocerin biosynthesis in Restharrow.

## 1.10 THE USE OF METABOLOMICS FOR BIOCHEMICAL CHARACTERISATION

Metabolomics can be described as ‘the systematic study of unique chemical fingerprints that specific cellular processes leave behind,’ (Daviss and Bennet, 2005). The metabolome is a collection of products from biological processes within cells, organs or organisms, which accumulate as either precursors or final products involved in cellular processes. ‘The isolation of

metabolites includes the lipid soluble chemicals normally found in plant membranes, polar chemicals found within the cell, basic and acidic ions, stable and non-stable oxidizing chemicals’.

Typically the tools used within metabolomics are nuclear magnetic resonance (NMR), Fourier transform infrared (FT-IR) spectrometry and mass spectrometry. More commonly, liquid chromatography (LC-MS) and gas chromatography mass – spectrometry (GC-MS), have been used to characterise the structures of several hundred chemicals within metabolomes, including, sugars, alcohols, organic acids, amino acids and a wide range of secondary metabolites, including those from the isoprenoid pathway (Haralampidis *et al.*, 2001).

### **1.11 METABOLOMICS IN PLANT BIOLOGY**

The use of metabolomics for studying the function, distribution and accumulation of metabolites in plants has become widely accepted as a standard technique within plant sciences. The current practice of metabolomics utilises diverse analytical platforms to detect a wide range of structurally different metabolites (Sumner *et al.*, 2003). The standard instrumental platforms include; GC-MS, LC-MS, Ultra-performance LC-MS, NMR and Capillary electrophoresis-MS (Haralmapidis *et al.*, 2001). These platforms can provide research groups with a qualitative and quantitative method for biochemical metabolite discovery of the metabolome from within an organism. Currently it is estimated there are approximately 1800 metabolites known to exist within plants, 900 of which have been fully characterised. In context it is thought that there are more than 10,000 metabolites existing within plants in nature.

Approaches towards plant metabolomics are highly dependent on temporal, physiological, environmental and spatial parameters, often making it difficult to directly compare results between experiments, species, tissue and growth stages (Farag *et al.*, 2008). The most sensitive modern analytical equipment available has the capacity to detect  $\sim 10^6$  number of analytes. In contrast the concentration range of metabolites thought to accumulate within

the plant cell is far greater and it is estimated at  $\sim 10^{12}$ . This suggests researchers only have the ability to detect  $\sim 50\%$  of metabolites available within the cell. It is estimated that an individual plant species can accumulate  $\sim 10,000$  metabolites whereas our current knowledge is less than  $\sim 20\%$  (Kushiro *et al.*, 1998). This however is considered as an acceptable value providing researchers with a way to take a real time ‘snap shot’ of the plant metabolome. There are many instances where plant metabolomics has been used to identify and characterise bio-molecules within the plant metabolome and listed below is a limited selection of groups who have reviewed plant metabolomics: (Kushira *et al.*, 1998; Fiehn *et al.*, 2000; Haralmapidis *et al.*, 2001; Sumner *et al.*, 2003; Corey *et al.*, 1997; Tohage *et al.*, 2005; Guy *et al.*, 2008).

### **1.12 AN INTRODUCTION TO THE TRANSCRIPTOME**

The transcriptome can be defined as the complete set of all RNA molecules including messenger RNA (mRNA) ribosomal RNA, (rRNA), transfer RNA (tRNA) and non-coding RNA within an organism. Unlike the genome, which is fixed for any group of cells the transcriptome varies considerably and can be influenced by other transcription factors and external stimuli. Transcriptomics refers to studying transcriptional changes using DNA microarray technology (Wan *et al.*, 2009).

Quantification of large numbers of (mRNA) transcripts using microarray technology can provide a detailed insight into cellular processes involved in the regulation of gene expression. Large-scale generation of gene regulation information was not possible until recently. Microarray technology (e.g. dot blots) was used in the mid 1980s (Augenlicht *et al.*, 1984, 1987), although did not become widely used until the mid 1990s, when complimentary deoxyribonucleic acid (cDNA) microarrays emerged (DeRisi *et al.*, 1996; Lockhart *et al.*, 1996). Microarray experiments are extremely powerful and provide information on a genome-wide scale. Further detail of transcriptional profiling technologies is outlined in chapter 6.1.

### 1.13 ARABIDOPSIS AS A MODEL SPECIES

*Arabidopsis thaliana* (Mouse ear cress) has been utilised as a plant model for studying mechanistic processes within triterpene biosynthetic pathways. Furthermore sterol biosynthesis has been studied to a level, not currently seen in any other plant organisms. A small plant metabolome, enables *A. thaliana* to be used for metabolite functional studies. *Arabidopsis thaliana* has a short generation time, requires little space to grow and produces large seed numbers, it is a diploid self pollinating organism. *Arabidopsis* was one of the first plant species to be fully sequenced in 2000 by the Arabidopsis Genome Initiative. 157 million bases and five chromosomes, are responsible for encoding the 27,000 genes and 35,000 proteins. (Gouw *et al.*, 2009). Much of the work has taken advantage in the fact that *A. thaliana* has a small genome and focused on functional characterization of plant genes by mutagenesis and T-DNA insertions, combined with gene expression and metabolomic studies through development. The advantages of having a fully sequenced genome have allowed the plant to be utilised in cross-hybridization expression studies using GeneChip® microarrays (Affymetrix 22k GeneChip®), (Hammond *et al.*, 2005). *A. thaliana* biosynthesizes several novel triterpenoids: namely maneral (2,3-oxidosqualene cyclase enzymes). Maneral is also biosynthesised in a distantly related plant species *Lactuca sativa* (Xiong *et al.*, 2006).

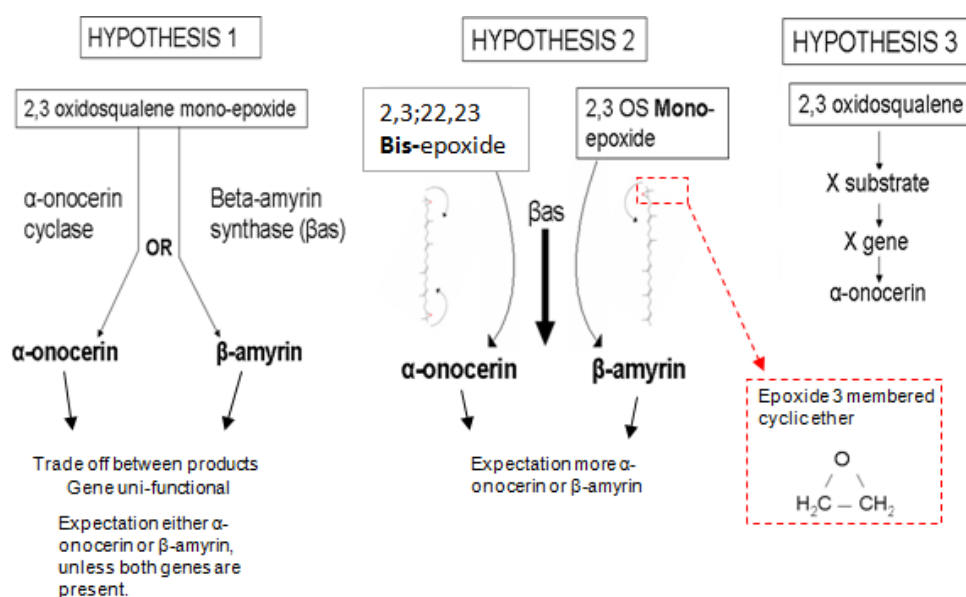
### 1.14 GENERAL PROJECT AIMS

Understanding the biosynthesis of  $\alpha$ -onocerin may provide us with a central example within plants of how triterpene diols are synthesized in nature. An investigation into whether a 2,3;22,23-dioxidosqualene is the primary precursor towards the synthesis of triterpene diols (such as  $\alpha$ -onocerin and serratene diol), may be of significance, as these compounds are distributed across a wide range of taxa. Evidence at the genomic, transcriptomic and metabolomic levels for bi-cyclisation events of the 2,3;22,23-dioxidosqualene, may raise further questions regarding the function of  $\alpha$ -onocerin and a plants ability to accumulate such compounds. The mechanism may also be mirrored in a number of other triterpene



biosynthetic pathways. Essentially this type of basic research would provide a framework for further functional characterization studies.

The main aim of this work was to investigate the biosynthetic route towards the cyclisation of  $\alpha$ -onocerin in Restharrow. A cross-disciplinary research approach was taken to test the three hypotheses surrounding  $\alpha$ -onocerin biosynthesis as outlined below:



**Figure 1.10** Three hypothesis suggest alternatives for  $\alpha$ -onocerin biosynthesis. Route 1; a direct cyclisation of 2,3-oxidosqualene, route 2) Cyclisation of a 2,3;22,23-dioxidosqualene, route 3) An alternative pathway.

1. A single cyclisation event from 2,3-oxidosqualene via a specific 2,3-oxidosqualene cyclase enzyme;  $\alpha$ -onocerin cyclase.
2. Cyclisation of a 23,22:2,3-dioxidosqualene, a product of squalene epoxidase - by either a  $\beta$ -amyrin synthase or a multi-functional enzyme (an oxidosqualene cyclase).
3. An alternative pathway – or a post cyclic modification, of either a 2,3-oxidosqualene or 23,22:2,3-dioxidosqualene product via a cytochrome P450.

Initially publically available databases were searched for genomic information regarding triterpene biosynthesis within the model crop *A. thaliana*. The aim was to determine which genes are involved in the *A. thaliana* triterpene pathway and identify regions, which may have similarities with triterpene biosynthetic pathway in Restharrow. The identification of likely transcriptional changes between tissue types or developmental stages, was also a key aim. This chapter also attempted to outline whether cross hybridization of *A. thaliana* was likely to occur on the ATH-1 GeneChip® Array. A gene list of AGI codes, probe sets and probe sequences were developed for use in cross species transcriptional profiling in chapter 6. The information in chapter 1 provided key information for subsequent, bioinformatic, metabolomic and transcriptomic experiments.

The aim of Chapter 4 was to outline which genes, enzymes and precursors are involved within the triterpenoid pathway in plants as a whole. The results were used for sequence alignments to develop genetic markers for key points within the triterpene pathway. An  $\alpha$ -onocerin cyclase has not been reported in the literature so it was important to deduce which enzymes were likely to be involved in  $\alpha$ -onocerin biosynthesis. Based on the literature and the results of Chapter 3 and 4, the enzymes involved in the cyclisation mechanism towards  $\alpha$ -onocerin biosynthesis are likely to be squalene epoxidases and oxidosqualene cyclases. The second aim was to identify the level of conservation at the amino acid level, between genes which occur in Restharrow and other closely related species such as *M. truncatula* and *A. thaliana*. Amino acid sequence analysis and three dimensional (3D) hypothetical structural modeling was utilised to outline structural differences between oxidosqualene cyclases.

Methods were developed using derivitisation and gas chromatography/mass spectrometry to extract, detect and quantify  $\alpha$ -onocerin and the relevant precursors in Restharrow. As a result of comparative metabolomics studies in Restharrows, it was deduced which species tissue type, time point where  $\alpha$ -onocerin and precursors accumulated. Information from Chapter 5 was use to focus transcriptional profiling studies.

Based on the metabolomic evidence the aim was to compliment the data with transcriptional information surrounding the expression genes involved in triterpene biosynthesis. The aim of the cross species *O. pusilla* and *O. spinosa* root and leaf transcriptional profiling experiment using Affymetrix ATH-1 GeneChip®, was to identify the transcriptional profile of a wide range of genes involved in the triterpenoid pathway. Initially the intensity values for perfect match (PM) oligonucleotide probes were analysed to identify any areas where cross hybridisation may occur between oligonucleotide probes. Following this, Common Data Format (CDF), file masks were developed at several intensity values for use with GeneSpring (version 11), for global transcriptional profiling. The transcripts of genes involved in triterpene biosynthesis were analysed and compared between species and tissue types.

## CHAPTER 2: MATERIALS AND METHODS

### 2.1 PLANT METHODS

#### 2.1.1 Plant materials

*Ononis spinosa* and *Ononis repens* were kindly supplied by British Wild Flower Plants and were used for initial derivitisation. Seeds were harvested and collected to be used at a later date. *O. spinosa* and *O. repens* seed were supplied by Naturescape British wild flowers seed company and used for all other experiments including cell free enzyme preps and a developmental growth series. *Glycyrrhiza glabra* and *Ononis rotundifolia* seeds were purchased from Chiltern seeds. *Ononis pusilla* seeds and *Ononis spinosa* plants were supplied from the previous project by Jane Kloda (Personal communication).

#### 2.1.2 Seed sterilisation

*Ononis* seeds were pre-treated with 10% (v/v) bleach (Parazone, Jeyes, Thetford, UK) for 10 minutes, then rinsed 3 times in sterile distilled water (SDW). The seeds were pre-saturated in 5 ml SDW at 35°C for 1 hour and scarified using a sterile scalpel before being germinated on 20 ml sucrose media (Fisher), pH 5.6, plus Gelrite (Sigma-Aldrich, Germany) gelling agent. This procedure was carried out under a laminar flow hood (Gelaire, Opera, Italy). The seeds were transferred to an environmentally controlled growth chamber (22°C; in continuous light ( $60 \mu\text{mol m}^{-2} \text{s}^{-1}$ ) where seeds were allowed to germinate until roots reached 10 mm in length.

#### 2.1.3 Plant growth

Seeds were pricked out after 7 days and potted into John Innes seed compost (Scotts, UK), 3 parts to 1 part perlite, (2.0-5.0 mm, Sinclair Lincoln, UK), used to improve drainage. Restharrowes were grown under a 16 hour light (24°C) and 8 hour (18°C) dark cycle. Light intensity was  $150 \mu\text{mol m}^{-2} \text{s}^{-1}$ . The relative humidity was ~70%, although it was not controlled.

## **2.2 NUCLEIC ACID METHODS**

### **2.2.1 DNA extraction from plant leaf and root tissue**

Plant genomic DNA was extracted from leaf and root tissue using a GenElute™ Plant Genomic DNA Miniprep kit (Sigma-Aldrich, Gillingham, UK). Initially 100 mg of plant tissue was ground to a fine powder using liquid nitrogen and a micro pestle and mortar. The ground tissue was placed into a 1.5 ml Eppendorf tube and reagents Lysis part A (350 µl) and Lysis part B (50 µl) were added and mixed by using a vortex. The solution was incubated for 10 minutes at 65°C, 400 µl of precipitation solution was added. In order to remove cellular debris, polysaccharides and excess protein, the sample was centrifuged for five minutes at 13,000 rpm. The remaining supernatant was transferred into a blue GenElute™ filtration column and placed inside a microtube (2 ml). The column was discarded after centrifuging for 1 minute at 13,000 rpm and 700 µl of Binding solution was added and mixed by inversion. 500 µl of Binding solution was added to the red GenElute column, which was then subsequently centrifuged at 13,000 rpm for 1 minute. The resulting supernatant was discarded and the column was placed in a new 2 ml Eppendorf tube. 500 µl of wash solution was added to the column. The column was centrifuged at 13,000 rpm for one minute. 50 µl of pre-warm (65°C) nuclease free water was added for the final elution step and centrifuged at 13,000 rpm for one minute.

### **2.2.2 RNA extraction MOBIO kit**

Total RNA was extracted according to the protocol supplied with MOBIO UltraClean™ (MOBIO, Carlsbad, USA), Plant RNA Extraction kit (Cambio Ltd. Cambridge, UK). 100 mg of plant tissue was homogenised using PMR1 with a micro-pestle in a 1.5 ml Eppendorf tube. The homogenate was vortexed and centrifuged at 10,000 rpm for 3 minutes. 600 µl of homogenate was added to 500 µl of PMR2 and 250 µl of PMR3 were added and mixed by inversion before incubating on ice for 5 minutes. Tubes were centrifuged at 10,000 rpm for 10 minutes. The supernatant was transferred to a new 2 ml Eppendorf tube and mixed with 800 µl of PMR4 by inversion. 650 µl of the sample was added to a spin filter and centrifuged at 10,000 rpm for 30 seconds. The spin filter was transferred into a new collection

tube and 500µl of PMR5 was added and the sample was centrifuged at 10,000 rpm for 90 seconds. 50 µl of PMR6 was added to the spin filter after to elute the RNA. The flow through liquid (purified RNA) was collected and stored at -80°C.

### **2.2.3 RNA extraction Qiagen RNeasy: high quality**

The Qiagen RNeasy Mini Kit (Qiagen, Crawley, UK) was used for purifying plant leaf and root tissue where high quality RNA was required for Affymetrix GeneChip® technologies (Affymetrix, Santa Clara, USA). As listed in the manufacturer's protocol, 100 mg of plant tissue was ground to a fine powder in liquid nitrogen using a pestle and mortar. 50 µl of RLT buffer was added and mixed by vortex. The extract was then pipetted onto a lilac QIAshredder column and centrifuged at 13,000 rpm for two minutes. 225 µl of absolute ethanol was added to the collected supernatant and mixed by inversion. 650 µl of extract was added to a new RNeasy Pink Spin Column and centrifuged at 13,000 rpm for 15 seconds. The flow through liquid was discarded before adding 700 µl of Buffer RW1. The column was then centrifuged at 13,000 rpm for 15 seconds. 500 µl Buffer RPE was then added to the RNeasy spin column and centrifuged at 13,000 rpm for 15 seconds and the flow-through liquid was discarded. 500 µl Buffer RPE was added to the RNeasy spin column and centrifuged for at 13,000 rpm for 2 minutes to further wash the spin column membrane. The RNeasy spin column was inserted into a new 1.5 ml Eppendorf tube and 50 µl RNase-free water was added directly to the spin column membrane and centrifuged for one minute at 10,000 rpm to elute the RNA.

### **2.2.4 RNA and DNA quantification**

A NanoDrop ND-1000 UV Spectrophotometer (Nanodrop, Wilmington, USA), was used to quantify the RNA and DNA concentration. The UV absorbance from 260/280nm from 1 µl of sample was measured to check the quality of DNA (1.8-2.0) and RNA (2.0 ±).

### **2.2.5 First strand cDNA synthesis**

Superscript III Reverse Transcriptase was used to reverse transcribe extracted RNA samples. The following was mixed and then incubated at 65°C for five minutes: 8 µl of supplied RNase free water, 10 µl of RNA, 1.0 µl of 10 mM Deoxyribonucleotide triphosphate (10mM dNTP mix) (NEB) and 1.0 µl of 50 mM Oligo dT<sub>18</sub> (Invitrogen). 1 µl of 0.1 M dTT, 1 µl of RNase out, 4 µl of first strand buffer and 1 µl of Superscript Reverse Transcriptase were added. The reaction mixture was heated for 60 minutes at 50°C and then inactivated at 70°C for 15 minutes.

### **2.2.6 Reverse Transcription**

A reverse transcription reaction was set up by adding 2 µl RNA (1 µg), 4 µl dNTP Mix (concentration per 5 mM dNTP, 2 µl of 3'RACE adapter, 2 µl of RT 10x Buffer, 1 µl RNase Inhibitor, 1 µl M-MLV Reverse Transcriptase and 8 µl Nuclease-free water. The reaction mixture was carefully stirred and centrifuged at 13,000 rpm for 15 seconds. The mixture was incubated at 42°C for 60 minutes and stored on ice for use as a RT-Reaction PCR template.

### **2.2.7 Polymerase Chain Reaction (PCR) methods**

Gradient PCR using a gradient PCR Express thermal cycler was used to optimise the annealing temperature of oligonucleotides. The PCR reagents quantity and concentrations that were used are listed in table 2.1. A gradient PCR cycle was set up with an annealing temperature range from 50-65°C, using the following PCR mixture for a 20 µl reaction.

PCR master mix reactions were aliquoted into 96-well ABgene PCR plates (Fisher Scientific, Loughborough, UK). A PCR rubber sealing mat was securely fastened and the plate was centrifuged, prior to mixing reagents. The same PCR reaction mixture was used for general PCR, changing the thermal cycling parameters where required (described in table 2.2).

**Table 2.1** The table lists the PCR reaction mixture details, volume in  $\mu\text{l}$  and concentration. NEB (New England Biolabs, Hitchin, UK).

PCR reagent	Quantity ( $\mu\text{l}$ )
Sterile distilled water	12.8 $\mu\text{l}$
PCR buffer (NEB)	2 $\mu\text{l}$ of 10 x
Reverse primer (Invitrogen or MWG)	0.5 $\mu\text{l}$ of 20 $\mu\text{M}$
Reverse primer (Invitrogen or MWG)	0.5 $\mu\text{l}$ of 20 $\mu\text{M}$
dNTP's (NEB)	0.4 $\mu\text{l}$ of 10 mM
MgCl (NEB)	0.6 $\mu\text{l}$ (25 mM)
<i>Taq</i> polymerase (NEB)	0.2 $\mu\text{l}$ (0.5 units/ $\mu\text{l}$ )
DNA template (cDNA/gDNA/plasmid)	2 $\mu\text{l}$

Primers were designed using primer3 ([http://frodo.wi.mit.edu/cgi-bin/primer3/primer3\\_www.cgi](http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)) and ordered from Invitrogen (Paisley, UK) or Eurofins/MWG (London, UK).

**Table 2.2** PCR thermocycler program parameters. The PCR step, time (minutes) and temperature ( $^{\circ}\text{C}$ ) of each step and the number of cycles is listed. \*= 1 minute per 1000 nucleic acids within the amplification target.

PCR step	Time and Temperature	Number of cycles
Initial denaturation	94 $^{\circ}\text{C}$ for 3 minutes	X 1 cycle
Denaturation	94 $^{\circ}\text{C}$ for 5 minutes	X 35 cycles
Annealing	50-65 $^{\circ}\text{C}$ for 1 minute	
Extension	72 $^{\circ}\text{C}$ for 1 minute/1kb*	
Final Extension	72 $^{\circ}\text{C}$ for 7 minutes	X 1 cycle

### 2.2.8 Rapid Amplification of cDNA ends (RACE) PCR

A FirstChoice<sup>®</sup> RLM-RACE Kit purchased from Ambion (Ambion, Invitrogen, Paisley, UK) was used to amplify the 3' region of  $\beta$ -amyrin – like synthase from *O. spinosa*. A series of nested primers were designed for the amplification of the 3' UTR. A PCR reaction mixture was prepared on



ice in the following order; 1 µl RT-Reaction PCR template, 5 µl 10x PCR Buffer, 4 µl dNTP Mix, 2 µl 3' RACE Gene specific (GSP) outer primer, 2 µl 3' RACE Outer Primer (supplied), 0.5 µl thermostable DNA *Taq* polymerase and 35.5 µl Nuclease-free water. The above was mixed by pipetting and placed on a thermocycler using the parameters as listed in table 2.2

### **2.2.9 Nested PCR 2<sup>nd</sup> round**

The template from the first round RNA ligase mediated (RLM) RACE PCR was used in a further nested PCR reaction and was set up in an identical reaction except for the primers. The outer primers; 3'RACE Gene specific (GSP) outer primer and 3'RACE Outer Primer were substituted with Inner primers GSP Inner primer and 3' RACE Inner primer (supplied). A thermal cycler program was set up as in table 2.8 PCR products were visualised using gel electrophoresis.

### **2.2.10 Agarose Gel Electrophoresis**

Agarose gel electrophoresis was used to separate amplified DNA PCR products. Agarose (Invitrogen, Paisley, UK) powder was made up to a concentration of 1-2% (w/v) in 0.5 x Tris-borate EDTA (TBE) buffer (0.4 M Tris, 0.4 M Boric Acid, 10 mM EDTA, pH 8.0) and heated in a microwave until fully dissolved. The gel was allowed to cool and set after (~50°C) and 0.5 µg/ml (final concentration), of ethidium bromide was added. Gel electrophoresis was carried out using a Biorad power pack (Biorad, Hemel Hempstead, UK) at 70-100 V. After an hour the gel was removed from the Biorad gel tank and visualised using a Biorad U.V transilluminator and Geldoc software (Biorad, Hemel Hempstead, UK).

### **2.2.11 Agarose gel extraction of DNA molecules**

For DNA fragments which required sequencing, agarose gel extraction was completed using GelElute<sup>TM</sup> Gel extraction kit (Sigma-Aldrich, Gillingham, UK). DNA amplicons were excised using a sterile scalpel whilst being viewed under UV (340 nm) light in the Biorad UV trans-illuminator

(Biorad, Hemel Hempstead, UK). The GelElute™ manufactures protocol was used to purify DNA products.

#### **2.2.12 PCR amplicon purification**

A QIAquick PCR Purification kit (QIAGEN, Crawley, UK) was used to clean PCR products prior to sequencing. The PCR products were diluted to a one in five ratio with the supplied PB buffer and mixed by inversion. The supernatant was then processed according to the manufactures protocol; QIAquick Spin Column centrifuged at 13,000 rpm for one minute, a column wash with 750 µl of wash solution, centrifugation at 13,000 rpm for 2 minutes. The DNA was eluted with 100 µl of nuclease free water.

#### **2.2.13 DNA Sequencing**

Pure DNA or purified PCR products were sent for sequencing to GeneService, Cambridge and Nottingham, UK ([www.geneservice.co.uk](http://www.geneservice.co.uk)). Primers were either supplied by GeneService or sent directly with the DNA samples as required (1 pmol/µl for each reaction). Flanking oligonucleotides M13 Forward (5'-TGTAACGACGCGCCAGT-3') and M13 Reverse (5'-CAGGAAACAGCTATGACC-3'), were used to sequence fragments cloned into pGEM T-Easy vectors.

#### **2.2.14 Ligation of PCR products into pGEMT-easy vector**

The plasmid vector pGEMT-Easy (Promega, UK) was used for sub-cloning. The *Escherichia coli* strain DH5α (Invitrogen, UK) was used for general manipulation of recombinant plasmids. The bacterial strain was obtained from Plant & Crop Sciences Division, Nottingham University, UK. PCR products were ligated into a pGEM T-Easy vector system (Promega, UK). A 3:1 insert to vector ratio was used in a 10 µl reaction. 5 µl of ligation buffer, 0.5 µl of vector and 0.5 µl of T4 DNA ligase were added to a purified DNA template and mixed by pipetting. The mixture was incubated overnight at 4°C.

### **2.2.15 *E. coli* transformation**

*E. coli* chemically competent cells DH5 $\alpha$  (Kindly supplied by Plant and Crop Science' Genomics facility), were slowly thawed on ice for 30 minutes. 5  $\mu$ l of the ligation with 100  $\mu$ l of DH5 $\alpha$  chemically competent cells were mixed and incubated on ice for 20 minutes. The cells were heat shocked by incubating in a water bath at 42°C for 45 seconds and then placed on ice. 900  $\mu$ l of LB broth was added to the cells and they were placed in a shaking incubator (220 rpm) at 37°C 1 hour. The cells were centrifuged for 1 minute and re-suspended in 200  $\mu$ l of LB before spreading on LB agar plates. 40  $\mu$ l of X-gal (20 mg/ml) and 40  $\mu$ l of 100 mM IPTG were spread onto the LB agar plates and the plates were allowed to dry at 37°C for 30 minutes. This allowed the transformed colonies to be screened using blue/white selection for the presence of an insert. The inoculated plates were incubated overnight at 37°C in an inverted position and white colonies putatively containing inserts were screened using colony PCR the following day.

## **2.3 MICROARRAY METHODS**

### **2.3.1 Oligonucleotide probe set identification**

Existing microarray databases were used for investigating the transcriptional profile of the triterpene synthases listed in table 3.1. Online tools GENEVESTIGATOR and eFP browser was used to study level of expression in tissues and during development. The Bulk Data Retrieval/Microarray element tool on the TAIR web page was used to identify probe set ID numbers by entering the locus names in AGI format. The probe ID was subsequently entered into the Probe Match dialogue box on the Affymetrix, NetAffx web page (<http://www.affymetrix.com/analysis/index.affx>), to retrieve all 18 individual probe sequences of each probe set for the genes as derived from the triterpene pathway.

### **2.3.2 Oligonucleotide probe set cross hybridisation evaluation**

The resulting individual probe sequences from each of genes in table 3.1 were checked against the Genbank database for cross hybridisation using the tool (BLAST), (<http://www.ncbi.nlm.nih.gov/BLAST/>) from the NCBI web site.

### **2.3.3 Genomic DNA hybridisation; BioPrime® labelling**

The following procedures were carried out by the National Arabidopsis Stock Centre: From *O. spinosa* at the 12 week growth stage, genomic DNA was extracted and quantified as described in 2.2.1. For genomic DNA probe generation the Invitrogen BioPrime® DNA (Cat. No. 18094-011) labelling System, was used. 100 ng of gDNA was dissolved in dilute buffer in a microcentrifuge tube. On ice, 20 µl of 2.5 x Random Primer solution was added and heated for five minutes at 100°C in a water bath. The samples were placed on ice immediately and 5 µl of dNTP (10mM) mixture and sterile distilled water was added to a total volume of 49 µl. The tubes were mixed by inversion. One µl of Klenow fragment was then added before vortexing and centrifuging at 13,000 rpm for 30 seconds. The mixture was incubated at 37°C for 60 minutes. After incubation the solution was placed directly on ice and 5 µl of Stop Buffer was added and the resultant solution was stored at -20°C. This resulted in a 10-40 fold amplification of probes.

### **2.3.4 Hybridisation**

The following reaction mixture was prepared for the gDNA hybridisation procedure: 3 µl BSA (50 mg/ml), 2.7 µl Herring sperm (10 mg/ml), 5 µl B2 control, 15 µl 20 x Eukaryotic Control, 150 µl 2 x Hybridisation Buffer, 70.3 µl DEPC water. 54 µl of the BioPrime® labelled sample was added and vortexed to give a total sample volume of 300 µl.

### **2.3.5 Affymetrix GeneChip gDNA hybridisation procedure**

A Wet chip hybridisation was performed by adding 100 µl of dH<sub>2</sub>O, 100 µl of 2 x Buffer to the sample and incubated at 99°C for 5 minutes and then at 45°C for 5 minutes. The sample was then centrifuged at 1,500 rpm for 5 minutes to remove the wetting solution. A GeneChip ATH-121501

(Affymetrix, Santa Clara, USA) was eluted with 200 µl of sample and left to hybridise overnight incubating the chip at 45°C for 16 hours at 60 rpm in a shaking incubator.

### **2.3.6 Gene expression file creation**

The Gene Chip array was scanned on an Affymetrix G2500A GeneArray scanner and a cell intensity file (.cel) was generated using Microarray Analysis Suite (MAS Version 5.0; Affymetrix). This .cel file contained the gDNA hybridisation intensities between *O. spinosa* and *O. pusilla* genomic DNA fragments and all *A. thaliana* probes. Probe pairs were selected from the .cel file for further analysis using a X species parser script (v1.1), written in the Perl programming language (<http://www.perl.com>). The script used an algorithm to create probe mask files (.cdf) at a range of user defined intensities from 0-1000 intensity threshold for each tissue/species (<http://affymetrix.arabidopsis.info/xspecies>).

### **2.3.7 Genomic hybridisation intensity**

Robust Means Analysis was used to normalise gDNA .cel files using Genestat. Hybridisation intensities were analysed at both the perfect match and mismatch probes for 16 genes within the triterpene pathway. The signal values for each individual probe at both the PM and MM (1-11), was plotted for both species *O. spinosa* and *O. pusilla* gDNA (leaf) against the mean value from ATH-121501.

### **2.3.8 RNA hybridisation**

GeneSpring (GX11, Agilent Technologies, Stockport, UK) was the software used to analyse gene intensity values and identify differentially expressed genes. Robust Multichip Averages were calculated using RMA function.

## **2.4 BIOINFORMATICS**

### **2.4.1 Biochemical pathway gene characterisation**

The Plant Metabolic Network (PMN) and Kyoto Encyclopaedia of Genes and Genomes (KEGG), were searched for metabolites/enzymes, understood to be involved in the triterpene biosynthetic pathway. The biochemical

pathway models were ordered according to the pathway which produced them and then by the order they occurred in their individual pathway. Complex organics, such as C<sub>30</sub> molecules, were ordered from basic primary metabolites and their precursors, through intermediates to final products. At each stage in the pathway of interest gene annotations were available allowing prediction of the molecular process which produces such molecules. Online tools were used to search for closely related genes to triterpene biosynthesis.

#### 2.4.2 Use of genome databases for analysing the genome

Genome databases were searched using amino acid sequences, BLAST searches and sequence alignments were used to explore the genome organisation of triterpene biosynthesis.

**Table 2.3** Online bioinformatics tools were used as listed in the table. The web link and the organisation source origin are also listed.

Organization	Web link
The National Center for Biotechnology Information (NCBI)	<a href="http://www.ncbi.nlm.nih.gov">http://www.ncbi.nlm.nih.gov</a>
European Bioinformatics Institute (EBI)	<a href="http://www.ebi.ac.uk">http://www.ebi.ac.uk</a>
The Institute for Genomic Research (TIGR) Rice and the Arabidopsis databases	<a href="http://www.tigr.org/plantProjects.shtml">http://www.tigr.org/plantProjects.shtml</a>
The Arabidopsis Information Resource (TAIR).	<a href="http://www.arabidopsis.org">http://www.arabidopsis.org</a>
Salk Institute Genomic Analysis Laboratory	<a href="http://signal.salk.edu">http://signal.salk.edu</a>
The Nottingham Arabidopsis Stock Centre (NASC)	<a href="http://arabidopsis.info">http://arabidopsis.info</a>
Kyoto Encyclopaedia of Genes and Genomes (KEGG)	<a href="http://www.genome.jp/kegg">http://www.genome.jp/kegg</a>
Plant Metabolic Network	<a href="http://www.plantcyc.org">http://www.plantcyc.org</a>
UniProt Protein database	<a href="http://www.uniprot.org/help/uniprotkb">http://www.uniprot.org/help/uniprotkb</a>
PLEXdb (Plant expression database).	<a href="http://www.plexdb.org">http://www.plexdb.org</a>

## **2.5 ANALYTICAL CHEMICAL EXTRACTION METHODS**

### **2.5.1 Plant sample preparation**

Plant extracts were prepared using methods adapted from previous approaches (Barton and Overton, 1955; Rowan and Dean, 1972; Orhan *et al.*, 2003; Rollenger *et al.*, 2005). Plant tissue was subjected to a variety of drying treatments prior to extraction, either freeze dried, partially dried or fresh. Root tissue was harvested and soil was removed by washing with tap water. The roots were washed thoroughly using a hose and a root washer before being saturated for 10 minutes in 5% (v/v) bleach and rinsed thrice with sterile distilled water. All plant samples were weighed on a Precisa (Precisa, Milton Keynes, UK) 125a precision balance prior to extraction.

### **2.5.2 Propan-2-ol extraction from plant tissue**

As a starting point, an attempt was made to extract  $\alpha$ -onocerin from *Ononis spinosa* root material using the following method: Freeze dried tissue was ground using a pestle and mortar in liquid nitrogen. Using 10 ml chemical resistant containers, 5 ml of propan-2-ol (IPA) was added to one gram of ground plant tissue and then vortexed. After centrifuging for 5 minutes at 10,000 rpm, the IPA was aspirated and added to 5 ml dichloromethane (DCM,  $\text{CHCl}_2$ ). The solution was mixed by vortexing, and then centrifuged for 5 minutes at 10,000 rpm. The solution was aspirated and added to the retained IPA solution. 100  $\mu\text{l}$  of extract was added to 900  $\mu\text{l}$  DCM before injection. All samples were stored at  $-80^\circ\text{C}$ .

### **2.5.3 Dichloromethane extraction from plant tissue**

100 mg sample of fresh plant tissue was ground in a pestle and mortar under liquid nitrogen, before being added to a high pressure 5 ml vial (Anachem). 1 ml of 6% (w/v) KOH was added to cleave esters (Rowan 1971). The sample was then heated in a GC oven (Fisons) for one hour at  $100^\circ\text{C}$ . After cooling, 2 ml of dichloromethane (DCM) was added to the plant extract and vortexed. After settling the DCM layer was removed using a Gilson pipette and added to 2 ml sterile distilled water. The DCM was washed thrice, before inserting into a 2 ml vial ready for injection. All samples were stored at  $-80^\circ\text{C}$  to prevent degradation.

#### 2.5.4 Ethanolic extraction from plant tissue

Before the KOH step 1.5 ml of 99.9%<sup>+/-</sup> (v/v) ethanol was mixed with 100 mg plant sample. The sample was heated in a secure oven at 78°C for one hour. Ethanol was evaporated by applying N<sub>2</sub> under a fume hood. One ml of 6% (v/v) Potassium hydroxide (KOH) was added to cleave esters. The sample was then heated in a GC oven (Fisons) for one hour at 100°C. After cooling, 2 ml of dichloromethane (DCM) was added to plant extract and vortexed. After settling the DCM layer was removed using a Gilson pipette and added to 2 ml sterile distilled water. The DCM was washed thrice, before inserting into a 2 ml vial ready for injection. All samples were stored at -80°C to prevent degradation. Modifications to the DCM extraction protocol included adding an ethanolic incubation step as standard and increasing the concentration of KOH from 6% to 20% (v/v)

#### 2.5.5 Hexane/ diethyl ether extraction from plant tissue

The protocol was adapted in order to extract a wider range of non-saponifiable lipids such as squalene, squalene mono-epoxide and squalene bis-epoxide. Based on the KOW values calculated by Estimated Programs Interface (EPIsuit, Windows<sup>®</sup>), diethyl ether was less hydrophobic and absorbed water, allowing the less hydrophobic non-saponifiable lipids to be extracted.

Solvent indicates LOG KOW for hexane and diethyl ether calculated using the EPI suit (a greater value = less soluble in water).

Hexane	=	3.2880
Diethyl ether	=	1.05
Dichloromethane	=	1.340
Ethanol	=	- 0.1412
Methanol	=	- 0.6323
Squalene	=	14.12

The extraction protocol used was adapted from Kolesnikova *et al.*, (2006). 500 mg of fresh plant material was homogenised in liquid nitrogen. The cell extract was then saponified with 2 ml of 10% (w/v) KOH in 80% (v/v)

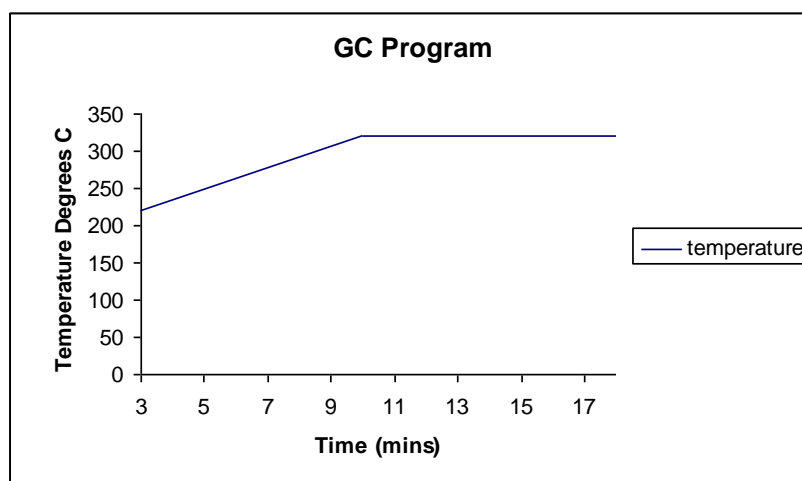


ethanol, after deoxygenating under technical grade N<sub>2</sub>. The suspension was incubated at 70°C for two hours. After cooling the reaction mixture was extracted three times with 2 ml Hexane/diethyl ether 10:1. The combined hexane/diethyl layers were washed twice with sterile distilled water. The upper layer non-saponifiable lipid (NSL) extract, was removed and placed into a new tube. The solvent was evaporated under a stream of N<sub>2</sub> to dryness and re-suspended in 1 ml of dichloromethane. Samples were stored at -20°C prior to direct injection.

## 2.6 GAS CHROMATOGRAPHY/ MASS SPECTROMETRY

### 2.6.1 Gas Chromatography using a Fisons 8000 series

Plant extracts were analysed using Fisons 8000 series (MD 800) Gas chromatography/Mass spectrometry or Thermo Scientific DSQ Trace GC Ultra (Thermo Fisher Scientific, Loughborough, UK). The Thermo Scientific DSQ Trace GC Ultra was a more advanced GC/EIMS made available part way through the project. GC/MS settings and experimental parameters were maintained.



**Figure 2.1** The temperature and time settings of the gas chromatography oven were used for GC-EIMS derivitisation. The temperature gradient increased from 200°C to 315°C within 10 minutes. A temperature of 315°C was held for a further 8 minutes.

When using the Fisons 8000 series all extracts were run through a non-polar 100% v/v), Dimethyl polysiloxane SGE 25 m x 0.22 mm x 0.25  $\mu$ m with a 1 m deactivated fused silica retention gap, to reduce residue build up.

The column was fitted following the manufactures guidelines. The carrier gas used was technical grade helium at a pressure of 30 PSI. GC settings: Column burn 360°C, Interface temperature 300°C, Source temperature 200°C, oven temperature 360°C, and Injection temperature 320°C, splitless (purge 2, split 2). Tuning parameters were set at peaks 69, 264, 502 and 614, using an auto tune function with a reference gas. GC oven temperature settings are shown in figure 2.4.

### 2.6.2 Gas chromatography/mass spectrometry: ThermoFisher DSQII

A Gas Chromatogram Mass Spectrophotometer (GC-MS) Thermo Scientific DSQ Trace GC Ultra (Thermo Fisher Scientific, Loughborough, UK) was used for GC-MS experiments. The GC-MS and auto sample files were set up using Xcalibur mass spectrophotometer software (Thermo Fisher Scientific, Loughborough, UK). Prior to analysing biological samples the Gas oven was heated to 300°C for a column burn and three blank samples of dichloromethane were injected to clear the column of any previous contamination.

**Table 2.4** The table below lists the injection parameters for the Thermo Scientific DSQ Trace GC Ultra. SSL (split/splitless) injector was subject to a gas pressure of 250 psi. The table lists the temperature, split flow rate, splitless time and the septum purge setting.

<b>Right SSL Injector: splitless, gas pressure 250 psi</b>	
<b>Temperature</b>	240°C
<b>Split flow</b>	20 ml/minute
<b>Splitless time</b>	1 minute
<b>Septum purge</b>	Constant

A Gas Chromatogram Mass Spectrophotometer Thermo Scientific DSQ Trace GC Ultra was used for auto-sampling injection of plant lipid extracts. 1 µl sample volume, 6 plunger strokes, injection depth standard with a pre injection dwell time were used. The same temperature gradient parameters were used as displayed in figure 2.1.

Phenominex Zebron (Phenominex, Macclesfield, UK) ZB-1, 30 meter length and retention gap (10m) 0.25 mm ID, 0.25 µm film thickness. Temperature limit -60°C minimum, 360°C maximum, was used with an isothermal reaction. 100% Dimethylpolysiloxane, part number 7HG-G001-11 S/N 121560.

**Table 2.5** The table below lists the operating mass spectrometry parameters including the; start time, detector gain, ion source charge, tune, scan rate and mass range, that were used to operate the GC-EIMS.

<b>Start time</b>	3.50 minutes
<b>Detector gain</b>	1 x 10 <sup>5</sup>
<b>Ion source</b>	200
<b>Tune</b>	~auto tune
<b>Scan rate</b>	amu/s 500, 1.2151/second
<b>Mass range</b>	50-450 m/e

### 2.6.3 Gas chromatography / mass spectrometry standardisation

Four different alkanes were initially used, described in 2.1.1 section, which were similar in size to  $\alpha$ -onocerin C22, C26, C28, C30 and dissolved in 1 µg / ml DCM (HPLC grade 99.9% Fisher, Loughborough, UK). C22 alkane was undetectable due to its low retention time. The aim was to find a compound with a similar retention time as  $\alpha$ -onocerin, which would be unaffected by the extraction protocol, and could be used for  $\alpha$ -onocerin quantification. Mass spectra results were directly compared with mass spectra from the literature; alkanes  $\alpha$ -onocerin (Rowan 1971),  $\beta$ -amyryn (Kajikawa *et al.*, 2005) and NIST MS, mass spectrum library (2.0). All

internal standards were added and vortexed prior to saponification of extracts at a concentration of 1mg/ml DCM (Sigma-Aldrich technical data 1997).

To avoid blocking the GC column and to allow accurate quantitation of  $\alpha$ -onocerin and the internal standard, all internal standards were injected at either 1/10, or 1/100 dilution.

#### 2.6.4 Calibration of gas chromatogram peak area values

Calibration curves were generated by adding known concentrations of the analyte being detected and the internal control. Peak area values could be measured and compared to the concentration of analyte in mg/g dry weight.

**Table 2.6** A C<sub>28</sub> standard was used to calibrate peak area of 100% m/e saturation against concentration in µg/ml. The concentration of internal C<sub>28</sub> standard, volume of stock solution, amount of dichloromethane added are listed. The mean peak area and calculated peak area per µg/ml was also listed. 1µg/ml = a peak area of 1758 Mass lynx units.

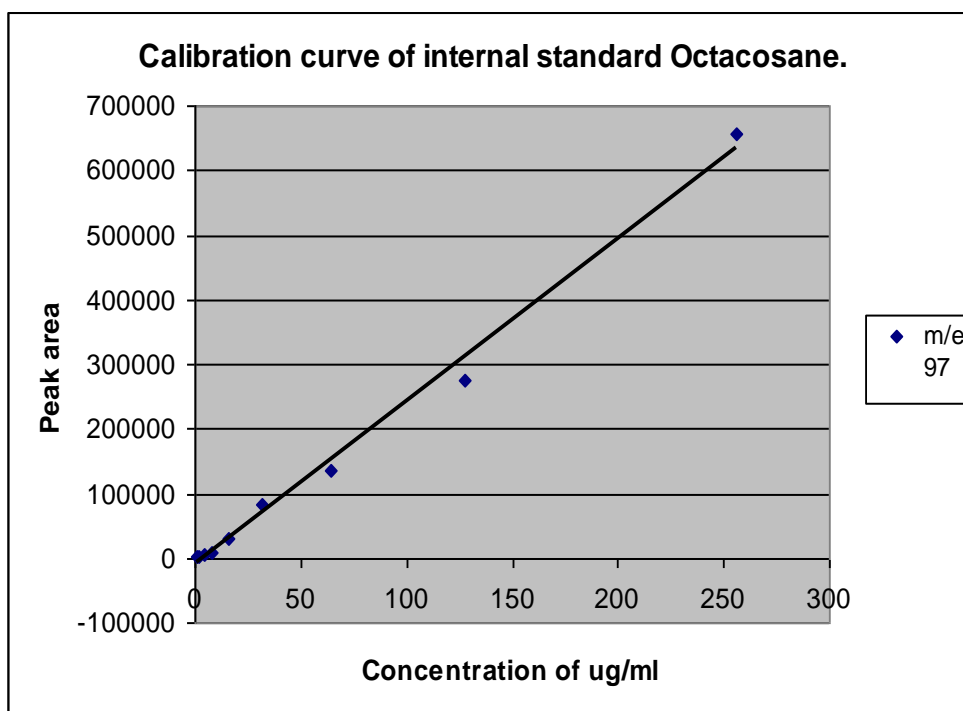
New concentration of C <sub>28</sub> standard (µg/ml)	Volume of stock added (µl)	DCM (µl)	Mean peak area	Peak area/µg/ml
1	0.5	499.5	751	327
2	1	499	1523	761.5
4	2	498	3945	986.25
8	4	496	8505	1063
16	8	492	30256	1891
32	16	484	81903	2559
64	32	468	134305	2098
128	64	436	273919	2139
256	128	372	657855	2569
				<b>1µg/ml=1758</b>

### 2.6.5 Mass spectral Selected Ion Remittance (SIR) detection

Selective Ion Remittance is a technique used in Mass spectrometry, once metabolites of interest have been derived. The technique has greater sensitivity compared to Total Ion Chromatography (TIC) by up to 100 times. The advantages of using this technique include, smaller data files, higher sensitivity and less, noise in chromatogram data. Selected ion remittance parameters were set according to the manufacture's protocol and were used where stated.

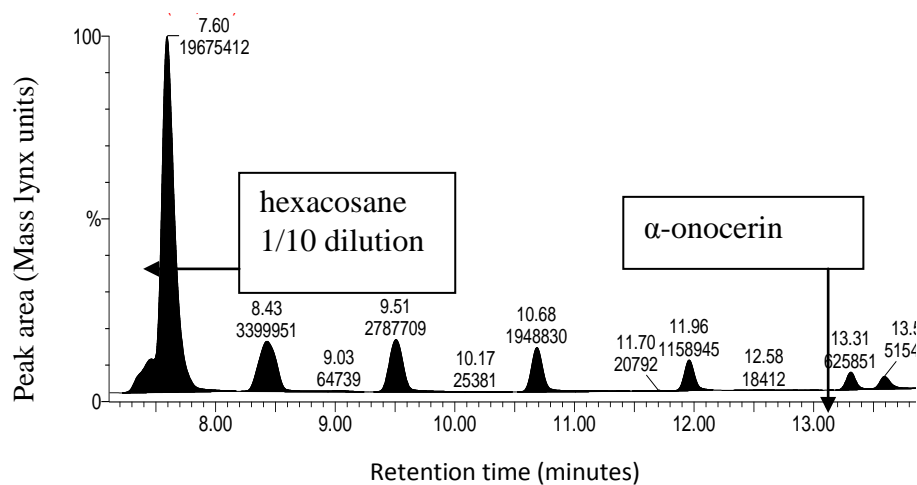
### 2.6.6 Quantification of non-saponifiable lipids

Selected ion remittance chromatograms were converted into TIC values to allow quantification. Peak area ratios were calculated using Mass Lynx v3.2 GC/MS analysis software.



**Figure 2.2** A calibration curve of internal standard octacosane, showing the concentration  $\mu\text{g/ml}$  to peak area (Mass lynx units) ratio. The molecular ion (m/e) used for quantification of the octacosane was 97, which produced a 100% saturation of the mass spectrum.

Comparisons between the Total Ion Chromatogram (TIC) peak area of a known internal standard and compounds of interest were calculated.



**Figure 2.3** A comparison of peak areas between hexacosane (RT 7.60 minutes) and  $\alpha$ -onocerin RT 13.59 minutes), used for quantification.

$1/38.1736 = 0.0261 \times 10$  (dilution Factor) = 0.261 mg of  $\alpha$ -onocerin in 100 mg x 10 = 2.61 mg/g  $\alpha$ -onocerin (dry weight).

## 2.6.7 Experimental design temperature time extraction efficiency

**Table 2.7** Two-factoral design using Design Expert programme 6.3, summarising experimental design parameters. Temperature in °C, and time in hours (hrs).

Sample run	Temperature (°C)	Time (hrs)
1	80	3
2	100	2
3	120	1
4	120	3
5	120	3
6	100	2
7	80	1
8	80	3
9	100	2
10	100	2
11	80	1
12	120	1

Once the compounds of interest were confirmed to be consistently detectable, extraction parameters were tested using experimental design software (Design Expert version 5, Stat-Ease, Minneapolis, USA). Temperature and time parameters of the potassium hydroxide incubation saponification step were tested using a Fisons GC oven.

## **2.7 EXPERIMENTS USING GC/MS SPECTROMETRY**

### **2.7.1 Non-saponifiable lipid fraction of Restharrow throughout development**

*O. spinosa* were germinated in MS0 media as previously described and transferred to soil 2 days after plant germination. Plants were grown in Levingtons M3 with horticultural silver sand in a 3:1 ratio, to ease root washing. Plants were placed in a growth room with a 22 hour photoperiod at 24°C.

1ml Ethanolic extracts were prepared and stored at -80°C until needed. At the same time point where possible, plant tissue was harvested, placed in liquid nitrogen and stored at -80°C for later use in qRT-PCR. All plant samples were biological triplicates and plant tissue was harvested from the same plants used for metabolomic data.

Internal standards C26 and C28 purchased from Sigma Aldrich (Gillingham, UK), were dissolved in analytical grade dichloromethane at a 1/10 dilution of 1 mg/ml and 0.25 mg/ml. The internal standards at these concentrations were used to calculate absolute quantities of metabolites involved in triterpene biosynthesis.

### **2.7.2 Non-saponifiable lipid fractions of Restharrow ecotypes**

Gas chromatography/mass spectrometry was developed used to analyse the chemical profile of non-saponifiable lipid fractions within several Restharrow plant species. Root and leaf extract from; *O. pusilla*, *O. repens*, *O. repens maritima*, *O. spinosa* and *O. rotundifolia*. *A. thaliana* and *G. Glabra* were included in the study. The same plant tissue that was used

for the cross species transcriptional profiling was used both at the 12 week development stage, where the metabolites of interest were most prevalent.

### **2.7.3 Cell free plant extract: plant tissue preparation**

Ten grams of fresh plant root material (12 week old *O. spinosa*), was homogenized under liquid nitrogen and added to 30 ml potassium phosphate buffer; (0.1 M K Phosphate; containing 2% (w/v) soluble polyvinyl pyrrolidone; 0.2 M Sucrose; 0.05 M ascorbate and 0.05 M cysteine (pH 7.2). The plant extract was passed through miracloth gauze (Calbiochem, Merk, Damstadt, Germany), whilst being agitated with a glass rod to remove cell debris. The remaining homogenate was collected and centrifuged at 5,000 rpm for 20 minutes in a 4°C centrifuge to remove any remaining cell debris. Each substrate; squalene, squalene monoepoxide (2,3-oxidosqualene), squalene bisepoxide (2,3-22,23-dioxidosqualene) and  $\alpha$ -onocerin, was prepared for inoculation by adding; 10  $\mu$ l tween 80, 10  $\mu$ l substrate, 10  $\mu$ l internal standard, 9970  $\mu$ l of DCM.

### **2.7.4 Cell free extract: substrate inoculation**

0.2 ml of substrate was added to 0.8 ml of plant supernatant. The solvents were removed in a fume hood under a stream of nitrogen using a sample concentrator at 40°C. A Pasteur pipette was used to chase the liquid around the bottom of the test tube, so that a thin film of substrate plus detergent was formed. 0.2 ml of potassium phosphate buffer was added and vortexed vigorously to emulsify completely. The boiled control preparation was incubated at 100°C for 30 minutes and then stored at 120°C. The non-boiled non-incubated control was placed at -20°C immediately to avoid any conversion of substrate by enzymes. The following samples were incubated over night at 25°C in a shaking incubator.

A comparison between inoculated cell free preparations and control (non-inoculated) extracts was made. Three test systems allowed a comparison between the accumulation of substrates using inactivated, incubated and natural control.



### **2.7.5 Extraction of cell free preparations: post incubation**

An equal volume (v/v, 200  $\mu$ l) of 10% KOH in 80% ethanol was added to the cell preparation vortexed and incubated at 60°C for one hour. After cooling the reaction mixture was extracted three times with 2 ml hexane/diethyl ether 10:1. The combined hexane/diethyl layers were washed twice with sterile distilled water. The upper layer NSL extract was removed and placed into a new tube. The solvent was evaporated under a stream of N<sub>2</sub> to dryness or in a fume hood and re-suspended in 1 ml of dichloromethane. Samples were stored at -20°C prior to direct injection. Chromatographic alumina was added to old ether to remove peroxide.

## **CHAPTER 3. THE USE OF PUBLICALLY AVAILABLE MICROARRAY DATABASES TO INVESTIGATE TRITERPENE SYNTHASE GENE EXPRESSION PATHWAYS IN *A. THALIANA***

### **3.1 INTRODUCTION**

#### **3.1.1 General introduction to publically available genomic resources**

Innovation in DNA-sequencing technology has led to a dramatic increase in the availability of DNA sequences. As a result the genome sequences from more than 300 bacteria and 30 archaeal genomes are available. In addition to this, an extensive library of nucleotide sequences from higher eukaryotes such as human (International Human Genome Sequencing Consortium, 2004), and plants *Arabidopsis thaliana* (The Arabidopsis Genome Initiative, 2000; Haas *et al.*, 2005), are available.

The genome and nucleotide sequence information being made publicly available, has revolutionised biological research and allowed the comparison of genetic information between genotypes. Known as comparative genetics, groups of genes are compared between organisms, species, or within species. This genetic information is often used to study evolutionary processes, environmental tolerance, and biosynthetic mechanisms of secondary metabolites (Hall *et al.*, 2002).

#### **3.1.2 An overview of microarray technologies**

This sudden explosion of available nucleic acid and amino acid sequence information has promoted the development of high-throughput comparative computational and microarray technologies. High throughput microarray based methods for gene expression monitoring were first described in 1995 (Schena *et al.*, 1995). The idea of gene expression analysis was first mentioned in the 1960's (Marmur and Lane, 1960) and blotting technology for expression came in the 1970's (Southern 1975). Micro-array technology has developed from these early technologies into a combination of genomics, transcriptomics, informatics, statistics and database searching.

Microarray technology has been applied to a wide range of applications across a diverse range of taxa within biology such as gene expression analysis (Schena, *et al.*, 1996), gene expression profiles for cancer cells (De Risi, *et al.*, 1996), single nucleotide polymorphism, splice-variant analysis, identification of unknown exons and analysis of DNA-protein interactions (De Risi, *et al.*, 1996).

Microarray *in situ* synthesis of oligonucleotides arranged in a high density format on glass slides, have become powerful tools for expression profiling in prokaryotic and eukaryotic organisms. Microarrays such as the ATH1 full genome array for *Arabidopsis thaliana* (23,750 genes) as developed by TIGR and collaborators, have been used to correlate gene expression changes with physiological and biochemical characteristics, to elucidate molecular biosynthetic mechanisms and responses to environmental stimuli (Hasegawa *et al.*, 2000).

The addition of thousands of publicly available arrays to repositories such as the Nottingham *Arabidopsis* Stock Centre (Craigon *et al.*, 2004), ArrayExpress at the European Bioinformatics Institute (EBI) and the Gene Expression Omnibus (GEO) at the National Centre for Biotechnology Information (NCBI), (Edgar *et al.*, 2002), have allowed microarray data to become readily accessible. As a result of these online databases tools have been developed on those sites and have allowed these large datasets to be mined.

### **3.1.3 Affymetix GeneChip technology**

There are two types of array; complimentary DNA (cDNA) arrays and Oligonucleotide arrays. Complimentary DNA arrays consist of probes from 150-X nucleotide base pairs (bp) in length, whereas synthetic probes consist of oligonucleotides of 24-80bp in length. Oligonucleotide microarrays have become the arrays of choice for use in gene expression studies. Oligonucleotide probes are fluorescently labelled during the cDNA synthesis stage from Ribonucleic acid (RNA) template, isolated from a

specific tissue to be analysed. The hybridized array is analysed using computational imaging, where the absolute gene expression level is measured. Each probe set (11 individual probes), is designed around a unique gene sequence and given an individual probe set identification number. In the case of Affymetix GeneChip technology, oligonucleotides are synthesised directly onto the chip through a photolithographic process, and are 25 nucleotides in length (Santa Clara, CA, USA; <http://www.affymetrix.com>). For the ATH1 GeneChip, each probe set consists of 11 Perfect Match (PM) and 11 Mismatch (MM; with a deliberate base Mismatch to the designed sequence at the 13th base) oligonucleotides. The perfect match versus Mismatch signal is used to test the hybridisation efficiency and can be used to normalise hybridisation signal strength (Liu *et al.*, 2002).

#### **3.1.4 Genevestigator as a resource for exploring microarray data**

Publically available genome arrays have led to several online interfaces developed to enable analysis and interpretation of microarray data. The Arabidopsis Information Resource (TAIR), AraCyc (Mueller *et al.*, 2003), MAPMAN, GENEVESTIGATOR (Zimmerman *et al.*, 2004), Arabidopsis eFP Browser, aGFP enable analysis of microarray data. To date several databases containing *A. thaliana* gene expression data are accessible, these are; NASCArrays (Craigon *et al.*, 2004), ArrayExpress, Stamford Microarray Database (SMD), Gene Expression Omnibus (GEO), (Edgar *et al.*, 2002).

In this chapter genes involved in the biosynthesis of triterpenes in *A. thaliana* were elucidated outlining AGI codes, oligonucleotide probe codes using GENEVESTIGATOR and PLEXdb. Pathway redundancy which may occur between *A. thaliana*, *M. truncatula* and *Restharrow* was identified using PLEXdb online tools, aiding the interpretation of data in Chapter 6. Cross hybridisation was evaluated by conducting BLAST searches with individual oligonucleotide probes. Individual tables are situated within the Appendix IV. Online gene atlas tools were utilised to build a map of signal values in order to; identify and analyse differences in signal values within

stages of triterpene biosynthesis: Squalene synthase (SQS), squalene epoxidase (SE), oxidosqualene cyclase (OSC), triterpene synthase (TTPS) and cycloartenol synthase and to build a gene expression atlas across the triterpene pathway between tissue types throughout development in *A. thaliana*. Along with providing background genetic information, the work in this chapter identified key homolog targets in specific tissues and developmental time points which may be targeted to investigate  $\alpha$ -onocerin cyclisation in *Restharrow*.

### 3.2 METHODS

In this chapter the expression profile of genes involved in triterpene biosynthesis in *A. thaliana* were studied using the online genome mining and expression profiling tools. GENEVESTIGATOR was utilised to develop tissue specific and developmental profiles for genes involved in the triterpene biosynthetic pathway (table 3.2). Gene expression discovery tools from GeneSpring version 10X were utilised to display gene expression information against a metabolomic map of triterpene and phytosterols biosynthesis.

Comparative expression profiling using Express Sequence Tags (ESTs) can be useful for outlining gene expression profiles, but caution should be taken when interpreting results as the presence of paralogues, multiple splice forms and orthologues with almost identical sequence may make data difficult to interpret (Schreiber *et al.*, 2009). Gene expression profiling can be considered as a method used to retrieve gene atlas expression information.

### 3.3 RESULTS

#### 3.3.1 AGI probes involved in triterpene biosynthesis in *Arabidopsis*

Initially, Arabidopsis Gene Identification (AGI) codes for genes involved in triterpene biosynthesis and phytosterol biosynthesis in *A. thaliana* were gathered from the literature. Plant Expression Database (PLEXdb) and TAIR (<http://www.tigr.org/plantProjects.shtml>) online tools were used to retrieve the probe set ID numbers. The structural details including the exon

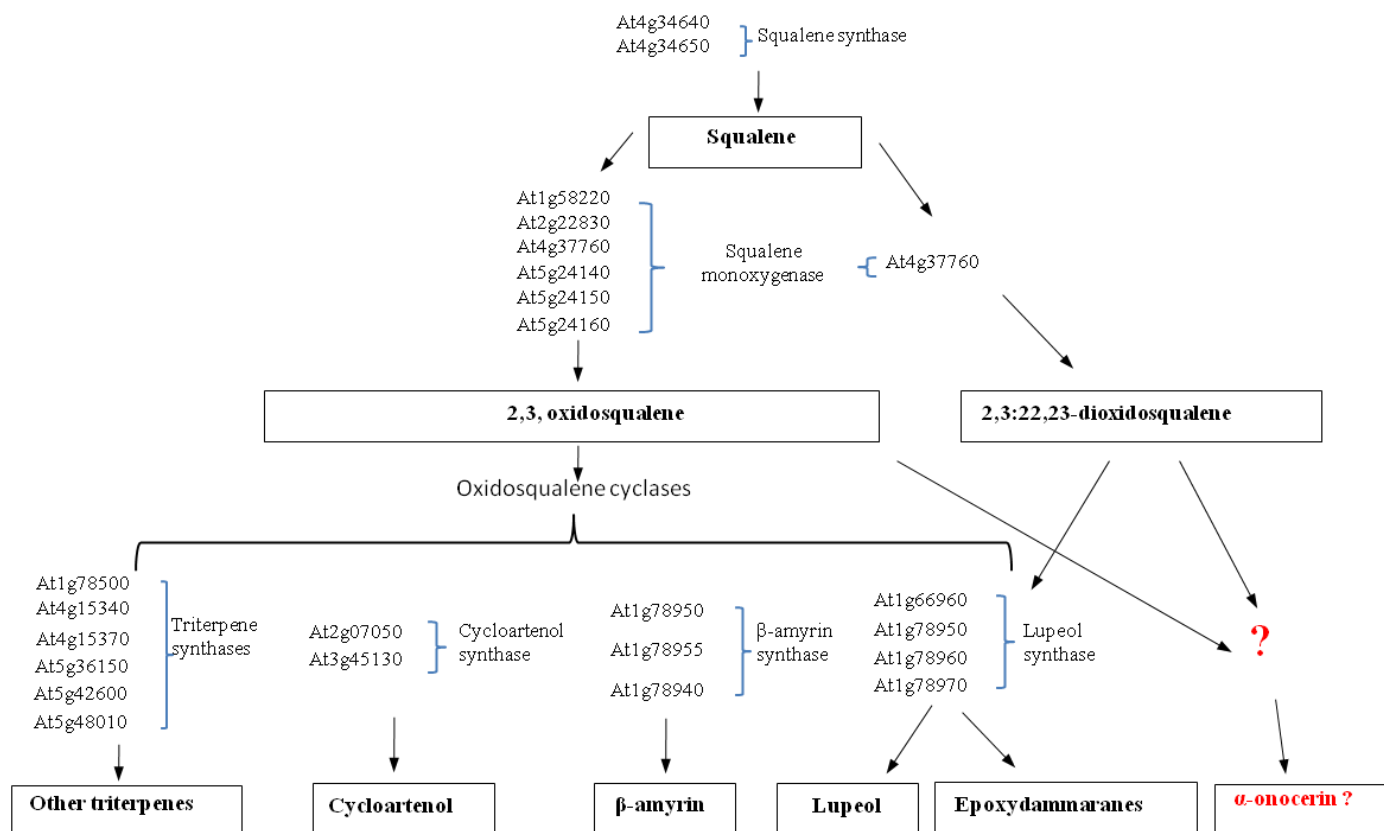
and intron information of individual genes involved in triterpene synthesis are listed in Appendix III.

**Table 3.1** The table lists secondary metabolites and genes involved in triterpenoid biosynthesis within *A. thaliana*. Listed are the abbreviations for the gene nomenclature and the Arabidopsis Gene Identification codes (AGI), (Auberg *et al.* 2002; Lange and Ghassemian 2003), used within GENEVESTIGATOR experiments. Probe set identification numbers are listed within the table used for cross hybridisation testing and cross species transcriptomics (Chapter 6). The number of exons within each gene is listed. ‘P’ Denotes a putative uncharacterised gene.

Biochemical product	Enzyme involved in biosynthesis	Gene name	AGI code	Probe set ID number	No. of Exons
Squalene	Squalene synthase	SQS1	At4g34640	253206_at	13
Squalene	Squalene synthase	SQS2	At4g34650	253254_at	13
Squalene epoxide	Squalene epoxidase	SQE1	At1g58440	245809_at	8
Squalene epoxide	Squalene epoxidase	SQE2	At2g22830	266831_at	6
Squalene epoxide	Squalene epoxidase	SQE3	At4g37760	253039_at	6
Squalene epoxide (P)	Squalene epoxidase	SQE4	At5g24140	249773_at	8
Squalene epoxide (P)	Squalene epoxidase	SQE5	At5g24150	249774_at	8
Squalene epoxide (P)	Squalene epoxidase	SQE6	At5g24160	249775_at	8
Lupeol	Lupeol synthase	LS1	At1g66960	255912_at	17
B-amyrin	$\beta$ -amyrin synthase	LS2	At1g78950	264138_at	17
Multifunctional	Lupeol synthase	LS3	At1g78960	264137_at	18
Lupeol	Lupeol synthase	LS4	At1g78970	264100_at	18
Seco-amyrin	$\beta$ -amyrin synthase	TTPS1	At1g78500	263123_at	14
Pentacyclic triterpene (arabidol)	Pentacyclic triterpene synthase	TTPS2	At4g15340	245258_at	14
Pentacyclic triterpene (Baruol)	Pentacyclic triterpene synthase	TTPS3	At4g15370	245553_at	14
Pentacyclic triterpene (P)	Pentacyclic triterpene synthase	TTPS4	At5g36150	249687_at	12
Maneral	Maneral synthase	TTPS5	At5g42600	249205_at	14
Pentacyclic triterpene (P)	Pentacyclic triterpene synthase	TTPS6	At5g48010	248729_at	13
Cycloartenol	Cycloartenol synthase	CAS1	At2g07050	266495_at	18
Cycloartenol	Cycloartenol synthase	CAS2	At3g45130	252611_at	17

Gene name, chromosome number, location on chromosome, a map of the gene model (TAIR), exon number and intron location are all outlined. The information presented directly showed that automatic gene calling is likely to be accurate for genes which may be linked. This is likely to be important information for genes yet to be functionally characterised situated adjacent to each other on the same chromosome. These include squalene epoxidases; At5g24140 (SQE4), At5g24150 (SQE5) and At5g24160 (SQE6).

As outlined in figure 3.1, triterpene biosynthesis is initiated by squalene synthase (At4g34640 and At4g34650). The subsequent biosynthetic step involves squalene epoxidase (squalene monooxygenase), At1g58220, At2g22830, At4g37760, At5g24140, At5g24150 and At5g24160 to cyclise 2,3-oxidosqualene or 2,3;22,23-dioxidosqualene as a bi-product. The stage where differentiation occurs between major triterpene products occurs within the oxidosqualene cyclases. Plant sterol biosynthesis is initiated with cyclartenol synthase; At2g07050 and At3g45130. Triterpene saponin biosynthesis is initiated by  $\beta$ -amyrin synthase; At1g78940, At1g78950 and At1g78955. Lupeol biosynthesis and the cyclisation of epoxydammaranes, is initiated lupeol synthase; At1g66960, At1g78950, At1g78960, At1g78970. In *Arabidopsis* triterpene synthases; At1g78500, At1g5340, At4g15370, At5g36150, At5g42600 and At5g48010, cyclising specific products present in *Arabidopsis*.



**Figure 3.1** A summary of the stages and genes involved in triterpene biosynthesis in *A. thaliana* deduced from Genevestigator, Plexdb and TAIR. Squalene synthase cyclising squalene At4g34640, At4g34650 ...



**Figure 3.1 continued...** squalene monooxygenase, At1g58220, At2g22830, At5g24140, At5g24150, At5g24160 cyclising 2,3-oxidosqualene and At4g37760 cyclising 2,3:22,23-dioxidosqualene. Oxidosqualene cyclases – triterpene synthases, At1g78500, At4g15340, At4g15370, At5g36150, At5g42600 and At5g48010 triterpenes. Cycloartenol synthase, At2g07050 and At3g45130, cyclising cycloartenol;  $\beta$ -amyrin synthase At1g78950, At1g78955, and At1g78940 cyclising  $\beta$ -amyrin. Lupeol synthase, At1g66960, At1g78950, At1g78960 and At1g78970 cyclising lupeol and epoxydammaranes.

### **3.3.2 Gene specificity of probe sequences from squalene synthase, squalene epoxidase and oxidosqualene cyclases from *Arabidopsis thaliana*.**

Individual oligonucleotide probe sequences were obtained using the NetAffx Affymetrix gene expression data discovery tool. Each individual oligonucleotide set (11 for each probe set) as listed in table 3.2, was assessed for sequence specificity using the BLASTN tool (TAIR). The individual sequences (perfect match individual probes) which are to be used in the cross species gene expression microarray analysis. Overall the probes showed high product specificity as indicated by a value of less than 1 E-value (Expect value). The E-value is a parameter which describes how closely a sequence matches a query sequence when subjected to a basic logical alignment tool (BLAST, NCBI). The closer the value is to 0 the greater the level of conservation between the sequences. The E-value takes into consideration the size of the database and the size of the query sequence (<http://www.ncbi.nlm.nih.gov>). For the purposes of expression analysis from *A. thaliana* as outlined in this chapter there is little chance of cross hybridisation of probes so they did not affect the microarray analysis.

**Table 3.2** Oxidosqualene cyclase (At1g78950) probe ID 264138\_at individual probe sequences 1-11 (NetAffx) and BLASTN (TAIR) sequence and score (E-value).

<b>Beta-amyrin synthase</b>	<b>AT1G78950</b>	<b>264138_at</b>
<b>Probe Sequence(5'-3')</b>	<b>AGI code</b>	<b>E value</b>
GAATGTACATCGTCTGCAATCCAAG	AT1G78950	7.00E-07
	AT5G66120	2.7
TTTTCAAGCAACTCTACCCTGATCA	AT1G78950	7.00E-07
	AT3G03360	0.17
GGACAACAGAGATCACCGCTTTCAT	AT1G78950	7.00E-07
	AT3G60400	0.67
GAACTGGGGCATTGCTTCACGTAC	AT1G78950	7.00E-07
	AT1G26170	0.67
TACGGTACATGGTTTGCTCTTGCA	AT1G78950	7.00E-07
	AT1G66960	0.67
GCTCTGCAGGCTTAGCAGCTGCGG	AT1G78950	7.00E-07
	AT4G37460	0.67
TGGGGAGAAAGCTACCTCTCTTGCT	AT1G78950	5.00E-07
	AT5G58250	0.47
GTGGTGCAAAGCTGCTTGGGCTTTAA	AT1G78950	5.00E-07
	AT5G04930	0.47
GATTTTCCTCAACAGCAAGCAACCG	AT1G78950	5.00E-07
	AT5G45150	0.47
GAAACATTCATCCGTTGTGGGCACT	AT1G78950	5.00E-07
	AT4G39753	1.9
AATATCGCGCGGAGTTTCGTTGCC	AT1G78950	5.00E-07
	AT4G00560	1.9

There are two squalene synthase homologues, six squalene epoxidase homologues, two lupeol synthase homologues, one  $\beta$ -amyrin synthase and one multi-functional lupeol synthase, six ATPEN triterpene synthases, two cycloartenol synthase homologues. Given the high relative levels of homology between a number of these genes (some of which appear to have arisen by gene duplication) it was expected that some cross hybridisation may have occurred between members of the probe pairs for the probe set to which they were designed and other homologues, being orthologues or paralogues. Evaluating potential cross-reaction between the Affymetrix oligonucleotides and non-target genes was important for comparative

expression analysis of *A. thaliana* and subsequent cross species analysis in *Restharrow*.

Each individual oligonucleotide probe sequence was subjected to the nucleotide BLAST function using Netaffyx tools available through the Affymetrix web site ([www.Affymetrix.com](http://www.Affymetrix.com)). All 20 genes currently known to be involved in triterpene biosynthesis were assessed for cross hybridisation based on the E-value as can be seen in table 3.3 as an example. An E-value of less than one indicated cross hybridisation was unlikely to occur. A value of greater than one indicated that cross hybridisation between oligonucleotide sequences may occur. Results for these experiments are present in the Appendix IV.

**Table 3.3** The summary table outlined below provides an example of how the E-value related to the number of matched individual probe oligonucleotides (outlined in the Appendix IV).

<b>E-value</b>	<b>Number of matched nucleotides</b>
<b>2e -04</b>	<b>25</b>
<b>0.011</b>	<b>22</b>
<b>0.043</b>	<b>21</b>
<b>0.66</b>	<b>19</b>
<b>2.6</b>	<b>18</b>
<b>10</b>	<b>17</b>
<b>41</b>	<b>16</b>
<b>162</b>	<b>11</b>

### 3.3.3 Comparative transcriptional profiling overview

GENEVESTIGATOR, which was used in this instance, provided a way to analyse gene expression profiles between development stages, and between tissue types tissue. GENEVESTIGATOR also provided an interface where gene expression was summarised as a metabolic pathway overview. Tools on other gene expression databases also offer some limited tool sets for

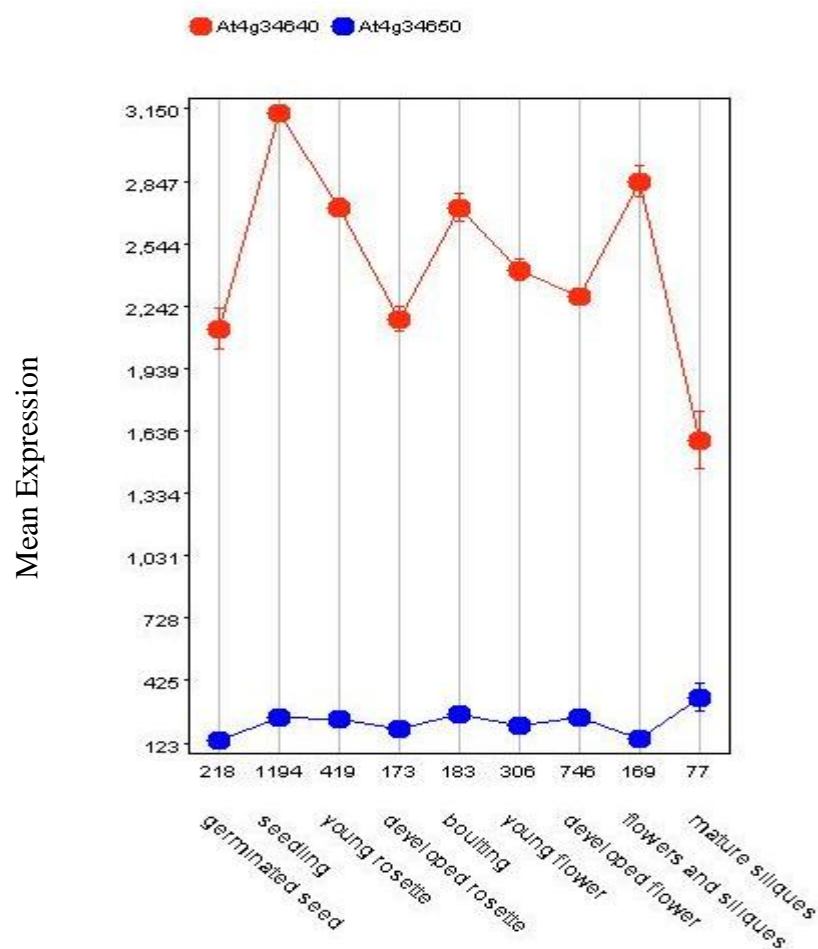
analysing gene expression between genes in a metabolic pathway context e.g. EGFP browser.

In the first instance the transcriptional profiles of genes within three specific stages of triterpene biosynthesis were defined using GENEVESTIGATOR Tools to plot the mean signal values between profiles. This allowed specific gene expression values to be outlined and interpreted in relation to function. Secondly Metaprofile tools were used to plot a gene atlas of triterpene biosynthesis transcriptional profiles across a pathway context. This allowed an overview of the triterpene to be developed and outlined which genes had significant roles to play in triterpene biosynthesis in *Restharrow*.

### **3.3.4 Comparative transcriptional profiling: squalene synthase**

During the squalene cyclisation stage of triterpene biosynthesis; At4g34640 squalene synthase is most highly expressed gene across all tissues within *A. thaliana*, during the seedling stage (3122 mean signal value), based on the Affymetrix datasets from GENEVESTIGATOR. At4g34640 showed a general high level of expression throughout development and had a higher expression level across all developmental stages than the At4g34650 squalene synthase homologue, which was expressed at a low level during all developmental stages, with the highest mean expression value was 336 during the mature silique stage.

As displayed in Figure 3.2, throughout development the overall variation of gene expression changed very little for squalene synthase 2 homologue At4g34650. As the conversion of farnesyl transphosphate to squalene is seen as the initial stage towards non-sterol and sterol biosynthesis, the results suggested that At4g34640 squalene synthase 1 was the primary gene responsible for regulating triterpene and phytosterols biosynthesis and changes in gene expression may affect the accumulation of subsequent precursors within the pathway.

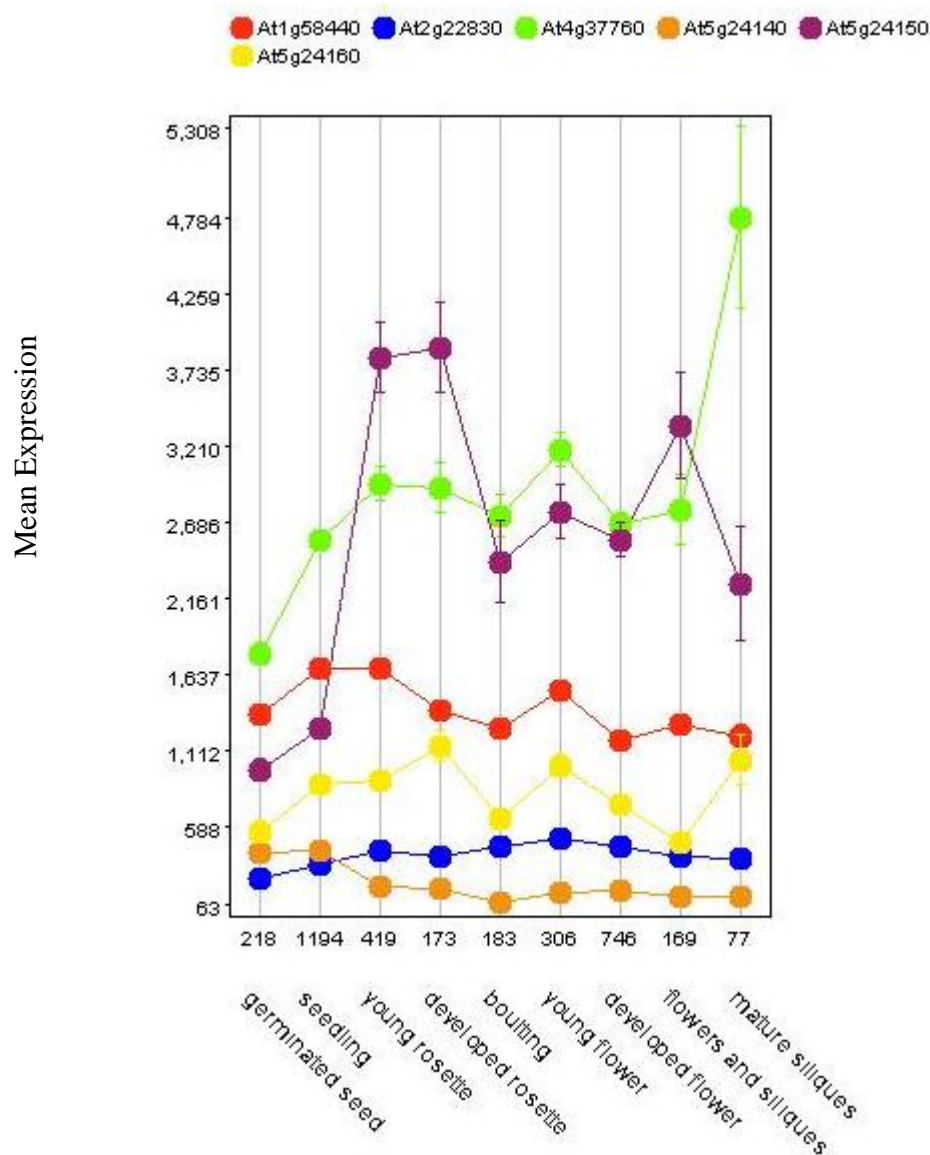


**Figure 3.2** The mean expression from all plant tissues combined; squalene synthetases (SQS), At4g24640 and At4g34650, were analysed at different developmental stages. The data was generated using high quality arrays, from within the Meta-profile analysis tool of GENEVESTIGATOR version 3 ([www.genevestigator.com](http://www.genevestigator.com)).

### 3.3.5 Comparative transcriptional profiling: squalene epoxidase

Six squalene epoxidase homologues were evaluated for gene expression levels throughout development, using the linear scale in Genevestigator shown in figure 3.3. Although there was a high level of conservation between sequences and evidence of gene duplication between squalene synthases, not all squalene epoxidases have been fully characterised in *A. thaliana*. At this stage in the biosynthetic pathway an epoxide group is added by an epoxidation reaction by a squalene epoxidase. In nature epoxidation usually occurs on the substrate asymmetrically in the active site

of squalene epoxidase, however it is suggested that 2,3;22,23-dioxidosqualene (bisepoxide) is produced as a by product as discussed in chapter 1 by the 2,3-oxidosqualene (monoepoxide) re-entering the active site to produce a symmetrical 2,3;22,23-dioxidosqualene.



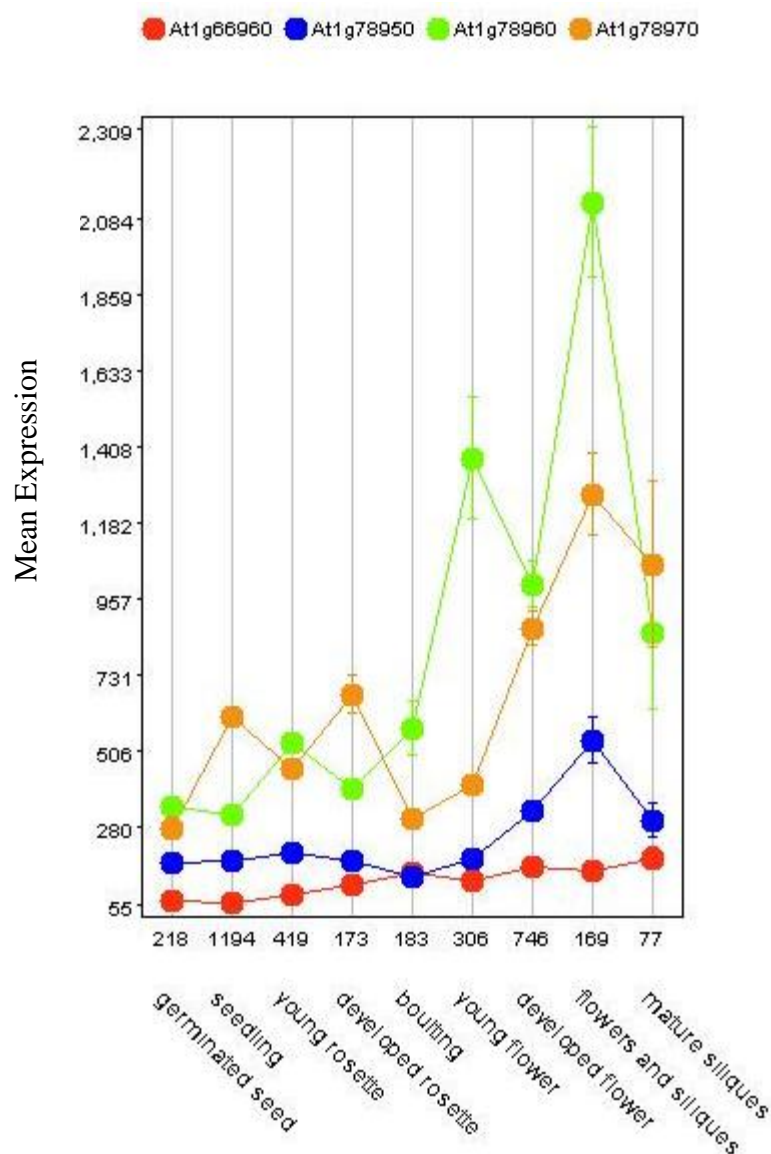
**Figure 3.3** The mean expression of all plant tissues combined from six squalene epoxidases (SQE): (At1g58440), At2g22830, At4g37760, At5g24140, At5g24150 and At5g24160 at different developmental stages was analysed. Data was generated using high quality arrays from the Meta-Profile analysis tool, GENEVESTIGATOR version 3 ([www.genevestigator.com](http://www.genevestigator.com)).

Gene expression patterns changed dramatically throughout development in two of the six homologues. The variation of expression values throughout development was limited for the remaining four homologues; At1g58440, At5g24160, At2g22830 and At5g24140. At5g24140 was expressed with a mean expression value of 407 during initial germination and development but decreased to a low level throughout the rest of plant development, suggesting this gene may be more important in early developmental stages than in the rest of the plants life cycle. At1g58440 mean expression decreased over time suggesting a role for this homologue in squalene epoxide biosynthesis during early development. At4g37760, increased in mean expression throughout development from 1750, during the seed germination stage to 4694 during plant senescence. This suggests that squalene epoxidase 3 may have a prominent role in squalene epoxide cyclisation throughout development. At5g24150 gene expression changed dramatically throughout development, when measured using GeneChip analysis. Initially two homologues, At4g37760 (1750) and At1g58440 (1345), were more highly expressed than At5g24150 (965) at the seed germination and seedling stages. Gene expression changed dramatically when the plant developed into the rosette stages to 3749 and 3817. During bolting and flower development At4g37760 expression (2684, 2631) was higher than At5g24150 (2364, 2705, 2631). The level of expression changed during mature silique production. Interestingly these changes in gene expression ratios were only seen in two of the six homologues. This may indicate that two of the gene homologues are primarily involved in controlling the synthesis of squalene epoxide which may alter in relation to changes in development, indicating channelling mechanisms within the pathway.

### **3.3.6 Comparative transcriptional profiling: oxidosqualene cyclases**

The meta profile tool from GENEVESTIGATOR was used to show the gene expression levels of four oxidosqualene cyclases (figure 3.4) during development. At1g78950 and At1g78960 have previously been characterised by yeast expression studies, as  $\beta$ -amyrin synthases and

At1g66960 and At1978970 as lupeol (and multifunctional triterpene synthase). At1g78960 was expressed at higher levels than At1g78950.

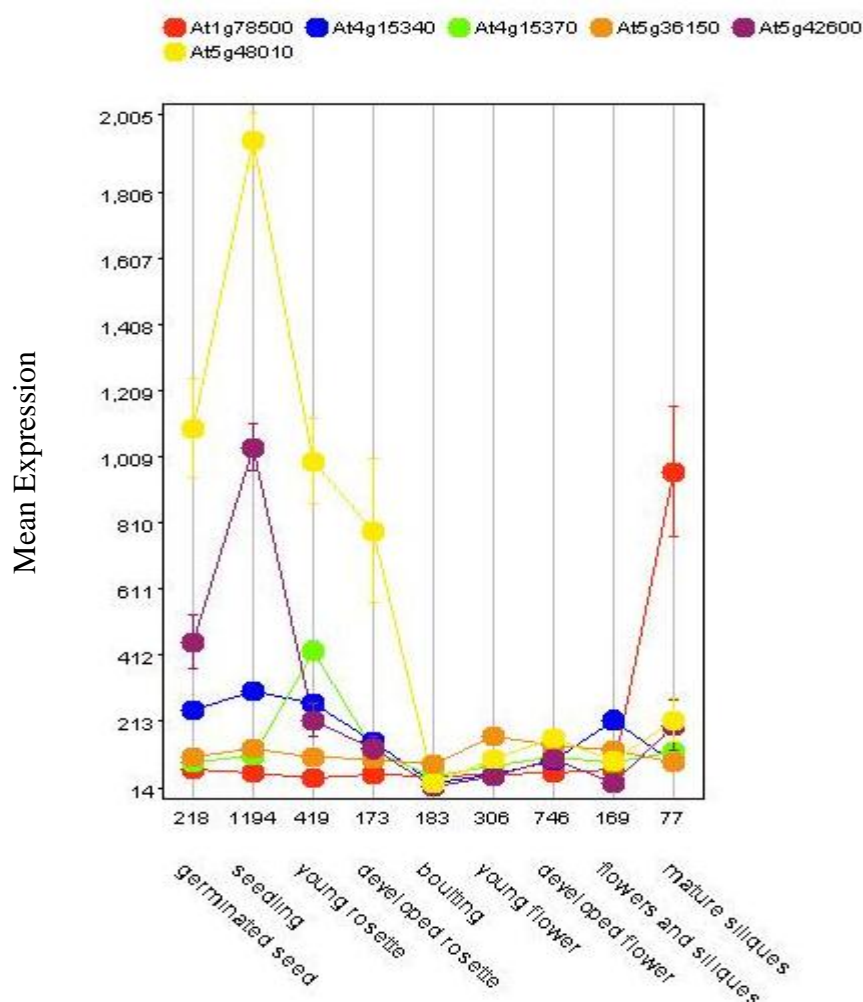


**Figure 3.4** The mean expression value from all plant tissues combined of oxidosqualene cyclases; At1g66960, At1g78950, At1g78960 and At1g78970 at different developmental stages was analysed. Data was generated using high quality arrays from the Meta-Profile analysis tool, GENEVESTIGATOR version 3 ([www.genevestigator.com](http://www.genevestigator.com)).

The expression of At1g78950 was at a consistent level throughout development except during the developed flower and silique stages, where the expression level increased to 556. Gene expression of At1g78960



increased throughout development. During the young flowering, developed flower and mature flower and silique stages, the gene expression increased dramatically from a mean expression level of 572 during bolting to 1,350 during flowering to 2,120 within the silique stages.



**Figure 3.5** The mean expression value from all plant tissues combined of six AtPEN oxidosqualene cyclases: At1g78500, At4g15370, At5g36150, At5g42600, At5g48010 at different developmental stages was analysed. Data was generated using high quality arrays from the Meta-Profile analysis tool, GENEVESTIGATOR version 3 ([www.genevestigator.com](http://www.genevestigator.com)).

The expression of lupeol synthase, At1g66960 was at a low level, with very little change throughout development. At1g78970 (multifunctional lupeol synthase) characterised by over expression studies in yeast, also showed an

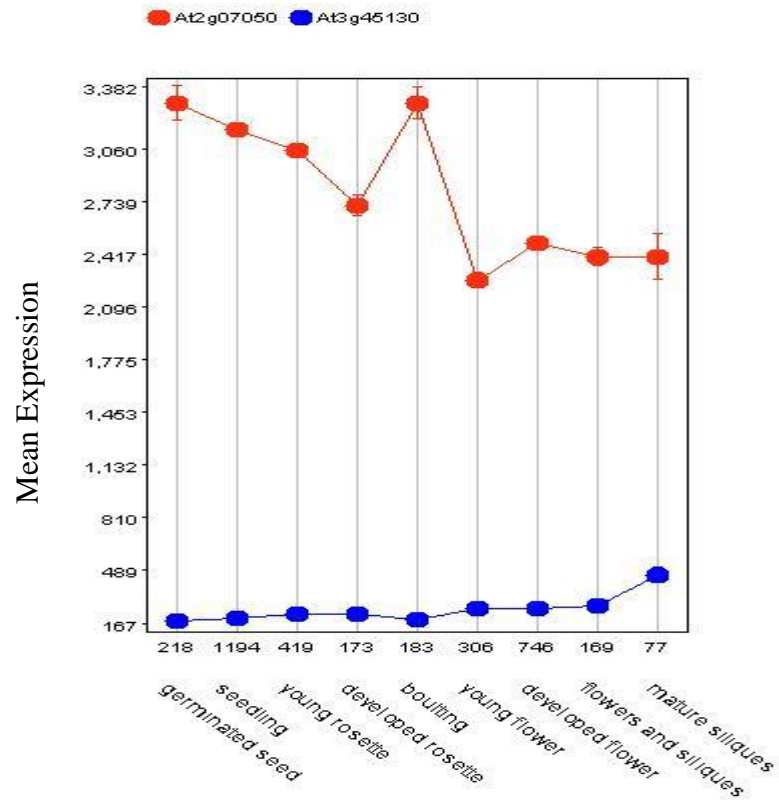
increased level of expression during later stages of development (Morita *et al.*, 2000).

As shown in figure 3.5, the recently characterised gene At1g78500 a  $\beta$ -amyrin synthase exhibited a low expression value across all arrays at all stages of development. The  $\beta$ -amyrin synthase which produces multiple products had a high expression value of 950 suggesting a functional role during the mature silique stage.

Gene clusters At4g15340 and At4g15370 (figure 3.5) showed similar low levels of expression throughout development and were most highly expressed at the early stages of development; At4g15340 at the seedling stages and At4g15370 at the young rosette stage. At5g36150 was expressed at a low level through all development stages. At5g42600 was expressed at its highest levels during the germinated seed and seedling development stages.

### **3.3.7 Comparative transcriptional profiling: cycloartenol synthases**

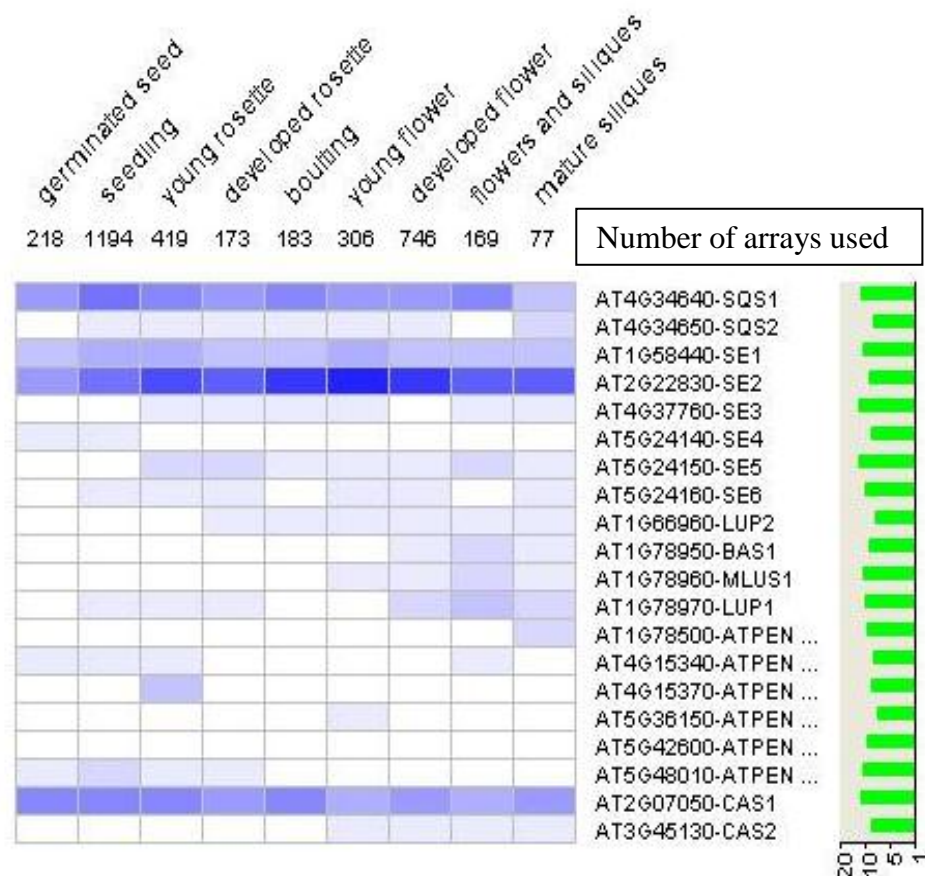
At2g07050, a cycloartenol synthase shown in figure 3.6, which catalyses the first stage towards sterol-triterpene biosynthesis was very highly expressed during all stages of development, the lowest expression level was 2201 during the young flower stage. A second homologue of cycloartenol synthase At3g45130 showed a significantly lower expression level, 171 during germination which gradually increased throughout development to 489 at the mature silique stage.



**Figure 3.6** The mean expression value from all plant tissues combined of two cycloartenol synthases: as before At2g07050 and At3g45130 at different developmental stages was analysed. Data was generated using high quality arrays from the Meta-Profile analysis tool, GENEVESTIGATOR version 3 ([www.genevestigator.com](http://www.genevestigator.com)).

### 3.3.8 Metaprofile analysis: through plant development

As outlined in figure 3.7 the changes in gene expression between developmental stages between gene sets were displayed



**Figure 3.7** A heat map generated by GENEVESTIGATOR: Meta-profile tool (<https://www.genevestigator.ethz.ch/>) showing a comparison of the mean expression value (Darkest colour top 1% of signal), of individual genes within the triterpene pathway throughout development. The colour intensity for each expression profile was normalised such that the highest intensity was defined as 100% (dark blue) and the absence of signal as 0% (white). The histogram adjacent to the heat map indicates the number of arrays used within the signal experiments used to generate the mean signal value.

**Squalene synthases:** At4g34640 (SQS1) squalene synthase 1 was consistently expressed during development, whereas At4g34650 (SQS2) squalene synthase 2 had a lower signal value throughout all development stages, although signal increased during the mature silique stage.

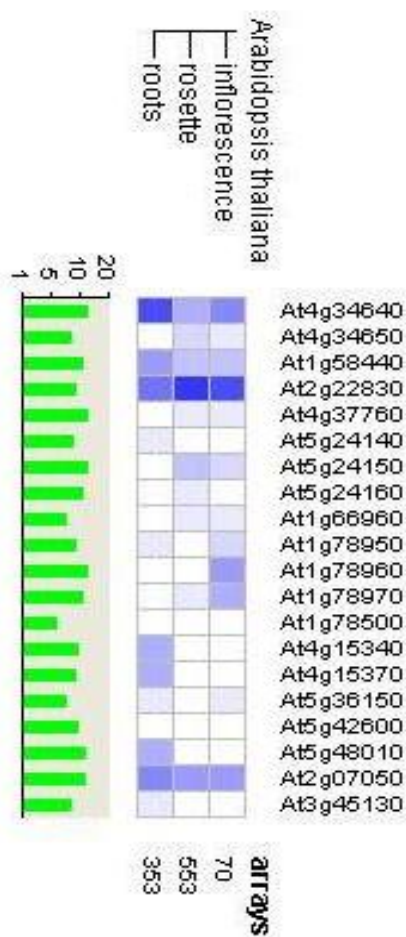
**Squalene epoxidases:** At2g22830 (SQE2) squalene epoxidase 2 produced a high signal intensity consistently throughout plant development. At1g58440 (SQE1) squalene epoxidase 1 produced a consistently high signal value second to At2g22830 (SQE2). At4g37760 (SQE3) produced no signal value during seed germination and the seedling development stages. The signal value was 0% during the flowering stage. In contrast, At5g24140 (SQE4) produced no signal during plant development but showed a weak signal value during the germinated seed and seedling stages. At5g24150 and At5g24160 produced similar signal values throughout plant development.

**Oxidosqualene cyclases:** At1g66960 (LUP2), lupeol synthase produced the most consistent signal throughout developmental stages. The signal value was lower than the threshold during seed germination, seedling development and young rosette stage. At1g78950 (BAS1)  $\beta$ -amyrin synthase and At1g78960 (MLUS1), multifunctional lupeol synthase both produced high signal values during flowering at the later stages of development. At1g78970, a multifunctional lupeol synthase (LUP1), produced a signal value during all plant developmental stages except the germinating seed stage. Within the group of uncharacterised putative oxidosqualene cyclases (ATPEN1-6). At4g15340 ATPEN2 was the only oxidosqualene cyclase to show a signal value during germinated seed, seedling and young rosette stages. At1g78500 (ATPEN1), triterpene synthase 1 showed an intense signal value during the mature silique stage only.

**Cycloartenol synthases:** At2g07050 (CAS1) cycloartenol synthase consistently produced a high signal value throughout *A. thaliana* development. At3g45130 (CAS2) cycloartenol synthase 2, produced a signal at a later stage in development, from young flowering onwards.

### 3.3.9 Metaprofile analysis: Organ specific comparative transcriptional profiling

The signal intensity for genes involved in triterpene biosynthesis varied between tissue types. To allow interpretation, the plant was split into three tissue types; root, rosette and inflorescence.



**Figure 3.8** Tissue specific gene expression (Darkest colour top 1% of signal) of triterpene synthases displayed as a heat map. A comparison between general tissue groups; inflorescence, rosette and roots. The histogram adjacent to the heat map indicates the number of arrays used to generate the mean signal value.

The signal value for At4g34640 (SQS1) was higher in the root than in the rosette or inflorescence. At1g34650 (SQS2), produced a signal in the rosette and inflorescence only. At1g58440 (SQE1) produced high signal values in

the root compared to other plant organs, whereas At2g22830 (SQE2) had a higher signal value in above ground plant tissues. Interestingly At4g37760 (SQE3) was expressed only in the rosette and inflorescence and in contrast At5g24140 (SQE4) produced the same signal intensity but not in the root. At5g24150 (SQE5) and At5g24160 (SQE6) produced no signal within the root. At1g66960 (LUS1) produced a signal in the above ground parts of the plant. At1g78950 (BAS1) showed signal within the root and inflorescence. At1g78960 ( $\beta$ -amyrin) was highly expressed in the inflorescence along with multifunctional lupeol synthase (At1g78970).

At1g78500 (ATPEN1) gave no signal. At1g5340 and At1g5370 pentacyclic triterpenes produced a signal within the root tissue only, as did At5g48010. At2g07050 produced high signal values across all tissue types, whereas At3g45130 cycloartenol synthase 2 produced signal values within root only.

### **3.4 DISCUSSION**

Tools from microarray browser databases provided a way to predict the putative function of gene sets. They can be used effectively for transcriptional profiling of metabolic pathways throughout development and between tissues. Expression profiles are useful for potential highlighting changes in pathway fluxes and outlining what pathway gene fluxes may be expected in other Angiosperms. The changes in gene expression between homologues of triterpene synthases raised biological and evolutionary questions.

#### **3.4.1 Expression of genes involved in the triterpene pathway throughout development**

Squalene synthase At4g34640 (SQS1) consistently produced a higher signal, than At4g34650 (SQS2) squalene synthase 2 during development. This may suggest that At4g34640 is the primary gene for controlling the biosynthesis of squalene synthase enzymes in Arabidopsis. The homologue At4g34650 showed an increased signal intensity during the mature silique stage suggesting an upregulation of the phytosterols and triterpene biosynthetic pathway during seed production in *A. thaliana*.

Squalene epoxidase 2, At2g22830 (SQE2) may be the primary gene encoding a squalene epoxidase during plant development based on evidence of high signal intensity throughout plant development. Squalene epoxidase 2 as discussed in chapter 3 has been reported to cyclise squalene epoxide into squalene bis-epoxide. Shuan *et al.* (2005) argue that a bis-epoxide intermediate may produce a more energy efficient way of cyclising pentacyclic triterpene, and may offer a feedback mechanism for regulating both phytosterol and triterpene saponin biosynthesis. The tissue specific metaprofile highlighted the fact that SQE1 had higher signal in root than in leaf, whereas SQE2 had a higher signal value in leaf.

Oxidosqualene cyclases; At1g66960 (LUP2), lupeol synthase At1g78970, a multifunctional lupeol synthase (LUP1) produced the most consistent signal strength throughout the available developmental stages. At1g78950 (BAS1)  $\beta$ -amyrin synthase and At1g78960 (MLUS1) multifunctional lupeol synthase had high signal values during flowering and at the later stages of development. At4g15340 ATPEN2 was the only oxidosqualene cyclase to have a clear expression value during; germinated seed, seedling and young rosette stages. At1g78500 (ATPEN1), triterpene synthase 1, produced an intense signal during the mature silique stage, suggesting specific functionality in this stage.

Cycloartenol synthase At2g07050 (CAS1) may be the primary gene responsible for encoding a cyclartenol synthase enzyme as a high signal level was seen consistently throughout *A. thaliana* development. At3g45130 (CAS2), cycloartenol synthase 2, produced a clear signal later in development, from the early flowering stage onwards.



### **3.4.2 Expression of genes involved in the triterpene pathway comparison between tissue types**

The expression data from GENEVESTIGATOR, as described in 3.3.4-3.3.7, indicated that the level of signal from gene homologues of squalene synthases, squalene epoxidases, and triterpene synthases varied between tissue types. In summary; At4g34640 (SQS1) At1g58440 (SQE1), At5g24140 (SQE4) At1g78950 (BAS1), At1g5340 and At1g5370, At5g48010 pentacyclic triterpenes and At3g45130 had high expression levels within the root. Triterpenoid  $\alpha$ -onocerin, primarily accumulates within roots in plants (Rowan and Dean 1972b) with this process likely to be controlled by Restharrow homologues of the genes outlined above..

### **3.4.3 Expression of genes involved in triterpene biosynthesis; a pathway overview**

The route towards triterpene biosynthesis in *A. thaliana* was explored using publically available databases. The *in silico* analysis of gene expression within *A. thaliana* outlined the complexity of studying specific functions of triterpene synthases within plants. The growth stage within plant development can dramatically affect the mean expression values of triterpene synthases, across multiple stages within the pathway, which may have an overall affect on the accumulation of secondary triterpenoids.

Evidence from GENEVESTIGATOR data suggested there is diverse variation between the Mean gene expression values across a wide range of tissue types, essentially leading to differences in the accumulation of triterpenes between different tissue types even within the same organs.

The genetic organisation of triterpenoid biosynthesis within the model crop *A. thaliana* genome, has been outlined. The high number of gene duplications present on the same chromosome at all stages within the pathway, outlined the challenges in developing complete oxidosqualene cyclase deletion lines which would be considered necessary for over expression and characterisation studies of triterpene synthases within *A. thaliana*. The information available offers an insight to how triterpene

biosynthesis may occur in higher plants where very little is known about the triterpene biosynthetic pathway. The data in this chapter has outlined specific areas within triterpenoid metabolism, which may affect the accumulation of triterpenes such as  $\alpha$ -onocerin in *Restharrow*.

Squalene synthase, A4g34640 was the primary gene involved in regulating the accumulation of precursors within the initial stages of triterpene biosynthesis. At2g22830 is most highly expressed across all tissue types and stages, suggesting it may be the most significant gene in controlling triterpene biosynthesis. Rasbery *et al.* (2007), suggest that SQE2 At2g22830 fails to cyclise 2,3;22,23-dioxidosqualene when over expressed in yeast. This may be significant when considering further studies in *Restharrow*, as in *A. thaliana* the availability for 2,3;22,23-dioxidosqualene may be limited, limiting the accumulation of epoxy-dammaranes.

At the oxidosqualene cyclase level, cycloartenol synthase At2g07050 was the most highly expressed gene across all tissue types and stages in *A. thaliana*. Oxidosqualene cyclases as a whole produced a higher signal value towards later developmental stages, which highlighted the importance for selecting the correct developmental stage in *Restharrow* for metabolomic and expression studies.

The pathway and expression data in this chapter can be used as a guideline as to what is likely to occur in other plant where little information is known, however it does by no means attempt to provide a comprehensive insight into triterpene biosynthesis in *Restharrow*. The information provides a starting point to identify key genes involved in triterpene biosynthesis in other closely related species, an indication of which growth stages and tissue should be considered when using comparative genomic, metabolomic and transcriptomic studies.

## **CHAPTER 4: THE USE OF BIOINFORMATICS TO INVESTIGATE TRITERPENE BIOSYNTHESIS IN HIGHER PLANTS**

### **4.1 INTRODUCTION**

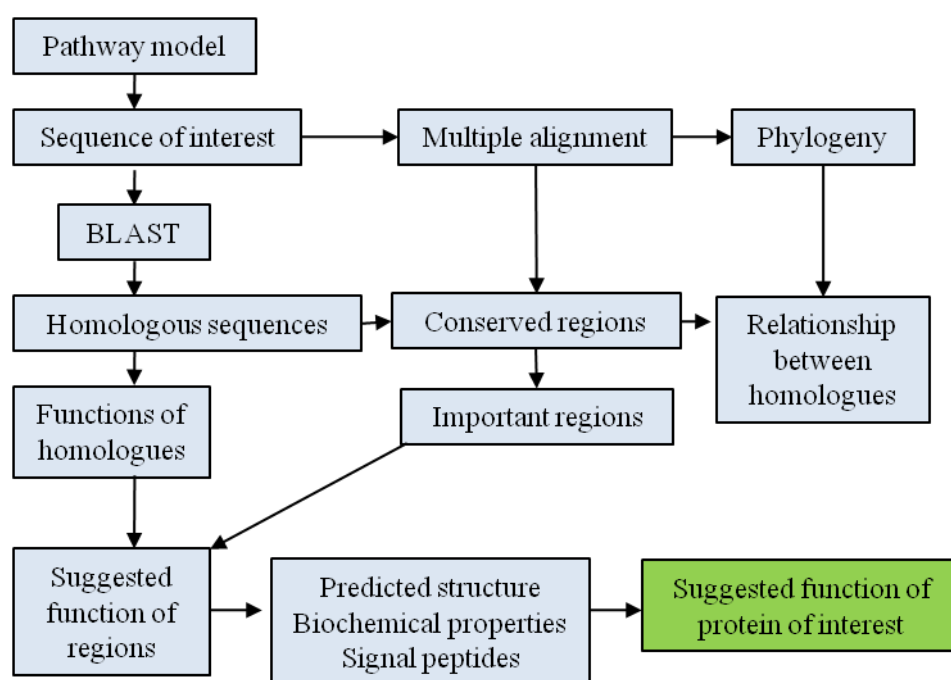
#### **4.1.1 General introduction to comparative bioinformatics**

Bioinformatics is the field of science in which biology, computer science and information technologies merge into a single discipline. Bioinformatics is comprised of three sub-categories; the development of new algorithms and statistics to determine the relationships between data sets; the analysis and interpretation of sequences of nucleotides, amino acids, proteins, their sub-domains and the development of tools which allow access to this information (Hogeweg and Searls, 2011).

A culmination of sequencing projects, with completion of the genomes of *Homo sapiens* in 2003 (*International Human Genome Sequencing Consortium*) and *Arabidopsis thaliana* in the year 2000 (Arabidopsis Genome Initiative 2000), led to the requirement of computational tools to interpret large data sets. By the year 2000 the GenBank repository of nucleic acid sequences contained 8,214,000 entries and the SWISS-PROT data-base contained 88,166 protein sequences. To date 500 organisms have been fully sequenced, ranging in coding content from 450 to over 100,000 genes. Bioinformatics technologies have been used as an interface and a tool to search and interpret available genetic data. Bioinformatics has allowed comparisons between organisms in terms of gene presence, structure, sequence and expression, allowing comparative genetic evolutionary studies.

Since the complete sequence was first published, bioinformatics has been used to interpret genetic information from *Arabidopsis thaliana*. As a result of developing tools for use in bioinformatics, a large resource of biological information is readily available. Each component of information can be

interlinked. By using chemical pathway blast searches with amino and nucleic acid sequences, multiple alignments and phylogenetic trees, biological information such as sequence, copy number, chromosomal location, location within the metabolic pathway and protein structure can be ascertained (figure 4.1). These tools allow researchers to ask important biological, physiochemical and evolutionary questions at the molecular and chemical levels, linking gene function with biological processes across organisms.



**Figure 4.1** A flow diagram outlining the process of bioinformatics whilst investigating the suggested gene function. (Neil Graham, unpublished, <http://plantsci.arabidopsis.info/pg/bio.html>).

The use of bioinformatics for phylogenetic studies and functional analysis follows a logical order. The gene or subset of genes of interest can be matched to similar genetic sequences using a 'Basic Logical Alignment Tool' (BLAST), as outlined on the National Centre for Biotechnology web page. A number of BLAST tools are available; nucleotide 'BLASTn', BLASTx searching nucleotide and translated nucleotide databases,

respectively, Tblastn, to search translated nucleotide databases using a translated query, to a straight forward protein search using 'BLASTP'.

In the case of identifying gene function from known sequences which have yet to be characterised, BLAST tools can be utilised to identify genes with high sequence similarity from other species. Sequences with highly conserved regions can subsequently be aligned with functional homologues to suggest the function of the protein of interest.

There is very little sequence information publicly available from plants with no commercial or obvious research value, or those plants considered unimportant as a food source. In this chapter the available bioinformatic resources were used to search and outline putative gene candidates involved in triterpene biosynthesis. By utilising genetic information from model crops (including *A. thaliana* and *Medicago truncatula*), valuable information may be gathered to develop tools to investigate  $\alpha$ -onocerin biosynthesis in Restharrow.

#### **4.1.2 Hypothetical pathway model for $\alpha$ -onocerin biosynthesis**

Bioinformatics and genome mining was utilised to generate a model of which genes are likely to be involved in triterpene biosynthesis in Restharrow. Hypotheses outlined in Chapter 1 were used to develop the key questions relating to  $\alpha$ -onocerin biosynthesis.

The following table outlines three possible ways  $\alpha$ -onocerin may be biosynthesised. Genome mining in other related species may offer an insight into which enzymes are likely to be present in Restharrow.

As shown in table 4.1, a squalene synthase, a squalene epoxidase (homology with a mono-epoxidase) and a number of oxidosqualene cyclases with a single function would indicate 2,3-oxidosqualene as an initial cyclisation product. The presence of a squalene synthase, a squalene epoxidase (showing homology to a characterised squalene bis-epoxidase) and a single oxidosqualene cyclase, would suggest a 2,3;22,23-dioxidosqualene as an

initial precursor. An absence of all three would indicate an alternative pathway.

**Table 4.1** Three possible hypotheses for  $\alpha$ -onocerin cyclisation within Restharrow. Hypothesis 1 assumes that the initial precursor is 2,3-oxidosqualene; Hypothesis 2 assumes that the initial precursor is 2,3;22,23-dioxidosqualene; Hypothesis 3 assumes an alternative precursor is involved in  $\alpha$ -onocerin synthesis. The presence or absence of expected metabolites from plant extracts may help identify the key enzyme involved in the biosynthetic mechanism.

Enzyme level	Hypothesis 1	Hypothesis 2	Hypothesis 3
	(Initial precursor 2,3-oxidosqualene)	(Initial precursor 2,3;22,23-dioxidosqualene)	(Alternative precursor)
Squalene synthase	Present	Present	Absent
Squalene epoxidase	Squalene epoxidase (squalene monooxygenase)	Squalene epoxidase with some homology with characterised squalene bisepoxidases	Absent
Oxidosqualene cyclase (OSC)	A number of different OSC's with a single function: $\beta$ -amyrin synthase, $\alpha$ -onocerin cyclase, or a cycloartenol synthase	A single gene copy having multiple products	Absent

## 4.2 METHODS

### 4.2.1 Profiling triterpenoid biosynthesis using online tools

The Plant Metabolic Network (PMN; <http://www.plantcyc.org>) and Kyoto Encyclopaedia of Genes and Genomes (KEGG; <http://www.genome.jp/kegg>), were searched for secondary metabolites reported in the literature to be involved in the triterpenoid biosynthetic pathway (Iturbe-ormetxe *et al.*, 2003; Xu *et al.*, 2004; Phillips *et al.*, 2006).

A limited amount of genetic information was available from the KEGG pathway tools outlining which enzymes were involved throughout secondary metabolite pathways. A 'molecular map' was deduced for triterpene biosynthesis in plants.

Aracyc (<http://pmn.plantcyc.org>) the plant metabolomic database was utilised to identify gene copy number involved in triterpene biosynthesis in closely related plant species. The Aracyc genomic database was searched using triterpenoid metabolites and precursors as queries. The same information was used to outline variation between enzymes and metabolites and where they occur within the triterpene pathway. The Aracyc genomic discovery tool, was utilised to identify gene clusters and possible gene duplications within the genome involved in triterpene biosynthesis. From this information a pathway map was generated to combine genetic information from within the triterpene pathway with a biochemical overview.

#### **4.2.2 Use of online tools for genome mining of triterpenoid pathway**

By using consensus sequences from closely related legume species aligned using ClustalP (<http://www.ebi.ac.uk>), degenerate primers were designed to be utilised as pathway markers in *Restharrow*. Subsequently genetic markers were designed from these sequences to amplify putative squalene synthases, squalene epoxidases and the  $\beta$ -amyrin synthase-like gene from *Restharrow*. An *Ononis* version of Actin was developed to be used as a control for RT-PCR.

The NCBI BLAST tools were utilised to assist in the identification of nucleic and amino acid sequences based on maximum score values of sequence identity, query coverage and E-values. Unrooted phylogenetic trees were produced using; NCBI Tree view, Accelrys gene software and phylogram online tools from the EBI web site within the sequence alignment output. NCBI was utilised to BLAST sequences from the *Restharrow* putative squalene cyclase, squalene epoxidase and  $\beta$ -amyrin

synthase (Kloda *et al.*, 2004) against fully characterised homologues within the model crop *A. thaliana*.

#### **4.2.3 Analysis of protein sequence structure**

EBI ClustalW was utilised, to conduct sequence alignments identifying regions of amino acid sequence variation and conservation. Expasy protein translation tools were utilised to translate nucleic acid sequences to amino acid sequences.

Once the active sites of triterpene synthases were identified from multiple alignments, SWISSProt (<http://www.uniprot.org/help/uniprotkb>) was used to generate PdB files to determine physical differences between the protein structure of *Pisum sativum* ( $\beta$ -amyrin synthase; PSY, BAA97558), multifunctional triterpene synthase (PSM, BAA97559) and *Ononis* putative  $\beta$ -amyrin synthase (Partial sequence, unpublished data, Mayes, 2004).

### **4.3 RESULTS**

#### **4.3.1 Triterpene profiling: use of online databases**

According to the KEGG plant triterpene pathway viewer, the first stage of triterpene biosynthesis consists of a conversion from a presqualene diphosphate to squalene by squalene synthase. Subsequently a squalene epoxidase, (mono-oxygenase), cyclases squalene into squalene epoxide (2,3-oxidosqualene). 2,3-Oxidosqualene acts as a precursor for further cyclisation by  $\beta$ -amyrin synthase or lupeol synthase. 2,3-Oxidosqualene has been reported by Matsuda *et al.*, (2000) as the initial precursor of phytosterol biosynthesis. Cycloartenol synthase has been reported to be a regulator of phytosterol production in cell walls (Haralampidis *et al.*, 2002).

#### **4.3.2 Comparative genome mining using Aracyc combined metabolome PMN and Genome tools**

An organism comparison was undertaken using an Aracyc tool from the NASC database (<http://arabidopsis.info/>). Aracyc combines genetic information publicly available into a pathway profile, the results of which are summarised in table 4.2.



**Table 4.2** An organism comparison as a result of information output from triterpene pathway analysis using the Aracyc tool from the NASC database (<http://arabidopsis.info/>). Gene accession names and the number of homologs of squalene synthase, squalene epoxidase and  $\beta$ -amyrin synthase for; *Medicago truncatula*, *Oryza sativa*, *Panax gensing*, *Solanum lycopersicum*, *Glycyrrhiza glabra*, *Avena strigosa*.

Step in pathway	Gene accession	Organism	No. of homologs
Squalene synthase	Mt1466.m00027	<i>Medicago truncatula</i>	1
	Os03g59040.1	<i>Oryza sativa</i>	2
	Os07g10130.1		
	BAA24289	<i>Panax gensing</i>	1
Squalene epoxidase	AJ430609	<i>Medicago truncatula</i>	2
	AJ430608		
	Pn-SE	<i>Panax notoginseng</i>	1
	SI-SGN-U315271	<i>Solanum lycopersicum</i>	4
	SI-SGN-U315272		
	SI-SGN-U317223		
	SI-SGN-U335		
$\beta$ -amyrin synthase	Mt-bas1-CAD23247	<i>Medicago truncatula</i>	1
	Gg-bAS1-BAA898	<i>Glycyrrhiza glabra</i>	1
	ASBAS1-CAC84558	<i>Avena strigosa</i>	1
	OSCPNY1-BAA33722	<i>Panax gensing</i>	2
	OSCPNY2-BAA33461		

#### 4.3.3 Arracyc: squalene synthase

In Table 4.2, outlined as the first stage towards isoprene biosynthesis was the enzyme squalene synthase which synthesises squalene. Two copies of squalene synthase in *Oryza sativa* (Poales; Os03g59040.1 and Os07g10130.1), there was a single copy of squalene synthase in *Medicago truncatula* (Fabales; Mt1466m00027) and *Panax gensing* (Asterids; PgPSS, BAA24289).

#### 4.3.4 Squalene epoxidase

*Panax notoginseng* (Apiales), (Table 4.2), has a single squalene epoxidase (Pn-SE), whereas in *Solanum lycopersicum* (Solanales) has four squalene epoxidases (SI-SGN-U315271, SI-SGN-U315272, SI-SGN-U317223, and

SI-SGN-U335064). There have been two squalene epoxidases reported in *Medicago truncatula* (AJ430609 and AJ430608).

#### 4.3.5 $\beta$ -amyrin synthase

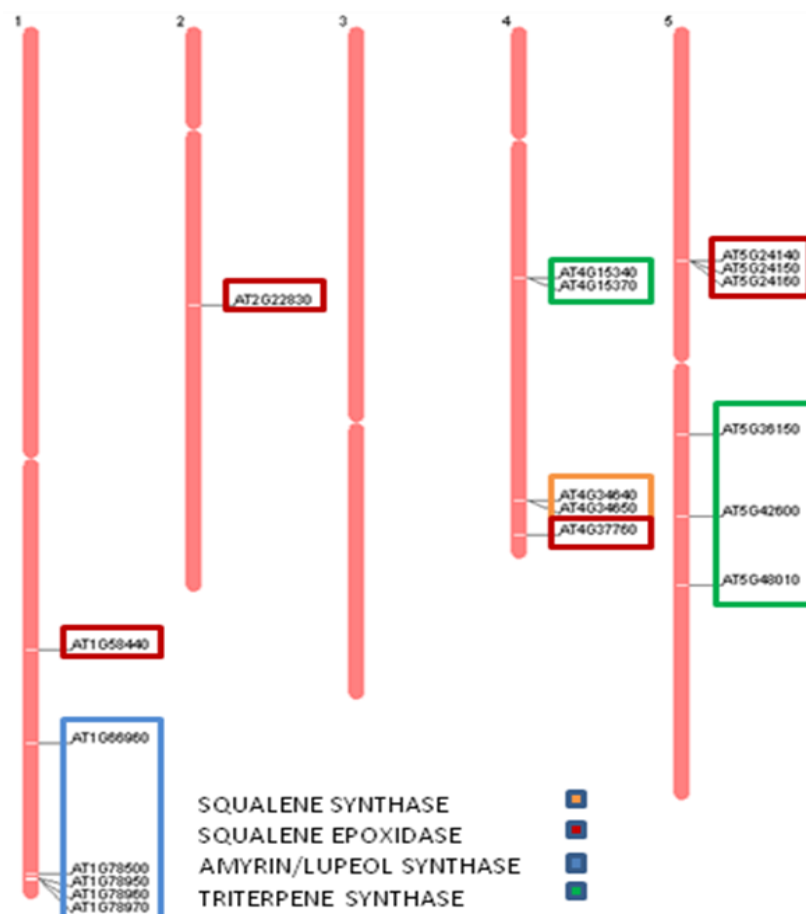
In *Glycyrrhiza glabra* (Fabales; Gg-bAS1-BAA898), (table 4.2), *Medicago truncatula* (Fabales; Mt-bas1-CAD23247), and *Avena strigosa* (Poales; ASBAS1-CAC84558), there was a single  $\beta$ -amyrin synthase gene. *Panax ginseng* (Apiaceae) had two homologues of  $\beta$ -amyrin synthase (Pg-OSCPNY1-BAA33722, BAA33461).

Information gathered from plant databases, Aracyc, PMN, BLAST searches, Affymetrix online tools and literature reviews (Lange *et al.*, 2002) allowed table 4.2 (section 4.3.2, page 93), to be developed outlining the genes involved in the triterpenoid pathway, their probe ID's and AGI codes within *A. thaliana*.

#### 4.3.6 Gene clustering of squalene epoxidases and $\beta$ -amyrin synthases

Gene chromosomal positions of triterpene synthases in *A. thaliana* were determined. The chromosomal maps obtained from AraCyc were subsequently annotated, to identify where gene duplications may have occurred.

Multiple copies of genes involved in the triterpene biosynthesis pathway were clustered at each stage. Two squalene cyclases; At4g34640 and At4g34650 are situated on chromosome 4. There are six copies of squalene epoxidase, three of which occur on separate chromosomes; At1g58440, (chromosome 1), At2g22830 (chromosome 2), At4g37760 (chromosome 4). Three putative squalene epoxidases; At5g24140, At5g24150, At5g24160 occurred as a gene cluster on chromosome 5.



**Figure 4.2** Chromosomal viewer output from Aracyc (<http://www.plantcyc.org>), showing the chromosomal location of each of the gene homologues within the triterpene pathway within *A. thaliana*.

In *A. thaliana* a significant volume of genetic information is readily available. Multiple copies of oxidosqualene cyclase enzymes in *A. thaliana* may offer alternative and perhaps more complex routes for triterpene biosynthesis, than in plants where a lower copy number of oxidosqualene cyclases can be observed. Complexity could also exist within the legume families, but with genomes which have not been characterised as thoroughly as *Arabidopsis*, this may not yet be fully recognised.

#### 4.3.7 Development of degenerate molecular markers for genes involved in triterpene biosynthesis

Amino acid sequences from *A. thaliana* for squalene synthase (SQS), squalene epoxidase (SQE1) and  $\beta$ -amyrin synthase (BAS), were used to

perform blast searches using NCBI TBLASTN (translated nucleotide blast using a translated protein query). The AGI codes for these genes are listed in table 3.1. In the example in Fig 4.3 below, the *A. thaliana* At1g78960 (multi-functional  $\beta$ -amyrin synthase) nucleic acid sequence was subjected to a tBLASTx (nucleotide translated query). A number of nucleotide sequences were extracted in FASTA format and aligned using ClustalW protein alignment (Accelrys Gene 2.0). Based on similarities in amino acid consensus sequences of *M. truncatula*, *O. sativa*, *P. sativum*, *A. thaliana* and *G. max*, degenerate oligonucleotides were designed. The BLAST searches and ClustalW (EBI) alignments were repeated for squalene synthase and squalene epoxidase.

Sequences producing significant alignments:

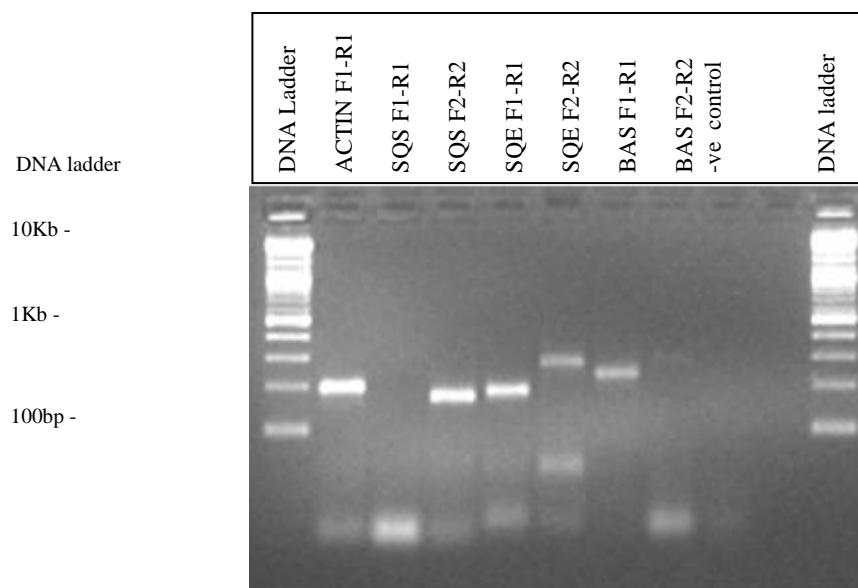
Accession	Description	Max score	Total score	Query coverage	E value
<a href="#">NP_178017.2</a>	ATLUP2; beta-amyrin synthase/lupeol synthase [Arabidopsis thaliana]	<a href="#">1610</a>	1610	92%	0.0
<a href="#">AAC17070.1</a>	Strong similarity to lupeol synthase gb U49919 from A. thaliana (see pentacyclic triterpene synthase [synthetic construct])	<a href="#">1593</a>	1593	92%	0.0
<a href="#">AAG41762.1</a>	pentacyclic triterpene synthase [synthetic construct]	<a href="#">1592</a>	1592	92%	0.0
<a href="#">XP_002887763.1</a>	ATLUP2 [Arabidopsis lyrata subsp. lyrata] >gb EFH64022.1  ATLL	<a href="#">1472</a>	1472	92%	0.0
<a href="#">AAC98864.1</a>	pentacyclic triterpene synthase [Arabidopsis thaliana]	<a href="#">1424</a>	1424	82%	0.0
<a href="#">XP_002887124.1</a>	ATLUP2 [Arabidopsis lyrata subsp. lyrata] >gb EFH63383.1  ATLL	<a href="#">1391</a>	1391	91%	0.0
<a href="#">NP_176868.1</a>	lupeol synthase, putative / 2,3-oxidosqualene-triterpenoid cyclase, LUP1 (LUPEOL SYNTHASE 1); beta-amyrin synthase/lupeol synthase [Arabidopsis thaliana]	<a href="#">1306</a>	1306	92%	0.0
<a href="#">NP_178018.1</a>	lupeol synthase [Arabidopsis thaliana]	<a href="#">1296</a>	1296	91%	0.0
<a href="#">AAD05032.1</a>	2,3-oxidosqualene-triterpenoid cyclase [Arabidopsis thaliana]	<a href="#">1294</a>	1294	91%	0.0
<a href="#">AAB94341.1</a>	hypothetical protein ARALYDRAFT_475791 [Arabidopsis lyrata subsp. lyrata]	<a href="#">1276</a>	1276	92%	0.0
<a href="#">XP_002887764.1</a>	hypothetical protein ARALYDRAFT_477065 [Arabidopsis lyrata subsp. lyrata]	<a href="#">1271</a>	1271	91%	0.0
<a href="#">BAF80443.1</a>	beta-amyrin synthase [Bruquiera gymnorhiza]	<a href="#">1230</a>	1230	91%	0.0
<a href="#">BAB83088.1</a>	beta-amyrin synthase [Betula platyphylla]	<a href="#">1230</a>	1230	91%	0.0
<a href="#">AAC17080.1</a>	Strong similarity to lupeol synthase gb U49919 and cycloartenol synthase [Arabidopsis thaliana]	<a href="#">1219</a>	2405	92%	0.0
<a href="#">NP_683508.1</a>	CAMS1 (Camelliol C synthase 1); beta-amyrin synthase [Arabidopsis thaliana]	<a href="#">1219</a>	1219	91%	0.0
<a href="#">XP_002889222.1</a>	hypothetical protein ARALYDRAFT_316793 [Arabidopsis lyrata subsp. lyrata]	<a href="#">1218</a>	2390	93%	0.0
<a href="#">CAD23247.1</a>	beta-amyrin synthase [Medicago truncatula]	<a href="#">1203</a>	1203	91%	0.0
<a href="#">XP_002330453.1</a>	predicted protein [Populus trichocarpa] >gb EEF08996.1  predicted beta-amyrin synthase [Medicago truncatula]	<a href="#">1201</a>	1201	91%	0.0
<a href="#">AAO33578.1</a>	beta-amyrin synthase [Medicago truncatula]	<a href="#">1200</a>	1200	91%	0.0
<a href="#">NP_178016.2</a>	beta-amyrin synthase, putative [Arabidopsis thaliana] >sp B6EXY6	<a href="#">1198</a>	1198	91%	0.0
<a href="#">BAE53429.1</a>	beta-amyrin synthase [Lotus japonicus]	<a href="#">1196</a>	1196	91%	0.0
<a href="#">ADK12003.1</a>	beta-amyrin synthase [Aralia elata]	<a href="#">1195</a>	1195	91%	0.0
<a href="#">ADK35123.1</a>	taraxerol synthase [Kalanchoe daigremontiana]	<a href="#">1194</a>	1194	91%	0.0
<a href="#">XP_002310349.1</a>	predicted protein [Populus trichocarpa] >gb EEE90799.1  predicted beta-amyrin synthase [Medicago truncatula]	<a href="#">1194</a>	1194	91%	0.0
<a href="#">BAA33461.1</a>	beta-Amyrin Synthase [Panax ginseng]	<a href="#">1192</a>	1192	91%	0.0
<a href="#">BAG82628.1</a>	beta-amyrin synthase [Arabidopsis thaliana]	<a href="#">1192</a>	1192	91%	0.0
<a href="#">BAA97558.1</a>	beta-amyrin synthase [Pisum sativum]	<a href="#">1192</a>	1192	91%	0.0

**Figure 4.3** A screen image from At1g78960 (multi-functional  $\beta$ -amyrin synthase from *A. thaliana*), translated nucleic acid sequence, is given as a central example of a TBLASTX, using the NCBI interface to search the Genbank database. The sequences with a high max score were used.

**Table 4.3** Degenerate primers designed to detect triterpene synthases in Restharrow. The name of the primer, location in sequence (\*if known) and degenerate sequence are shown in the table.

<b>ACTIN</b>			
	Primer name	Location of primer *	Sequence
1	Deg-Actin-F1		TGGHAACATYGTBYTSAGTGGTG
2	Deg-Actin-R1		GCCTTTGCAATCCACATCTG
3	Deg-Actin-F2		AGYAAYTGGGATGATATGGAG
4	Deg-Actin-R2		GACCHCTACCACAVTCAGTGTG
<b>Squalene synthase</b>			
5	DEG-SQCYC-F1	343 – 366	GCGRYNAGGMAMGCCGAGAAGCAG
6	DEG-SQCYC-R1	499 – 519	CGAGAVCTATGRCAACTCCTAC
7	DEG-SQCY-F2	1150 – 1175	CAGATMRTGCGRATTGGAACACTTG
8	DEG-SQCY-R2	1318 – 1341	CAACWGTTHTTTRCTAGGKTTACG
<b>Squalene epoxidase (Squalene monooxygenase)</b>			
9	DEG-SQE-F1		CTGCTCTTGCTTATACDCTTG
10	DEG-SQE-R1		CGAAAGTRTRCCCGCAAAAYAAG
11	DEG-SQE-F2		ATCTYGAATCCTTTTACACYCTKCG
12	DEG-SQE-R2		GGTRAAGAAACGVCAVCGMYA
<b>Oxidosqualene cyclase (<math>\beta</math>-amyrin synthase – like)</b>			
13	DEG-BAS-F1	385 – 404	CARACATGGGARTWTGATCCWG
14	DEG-BAS-R1	646 – 664	CTRCCRGTAACCGGMCGASTTT
15	DEG-BAS-F2	1742-1766	AAGAYCAYGGATGGCAAGTTTCTG
16	DEG-BAS-R2	2150-2171	CCKCAAACGAAGTGMSTRCCA

Reverse transcription-polymerase chain reaction (RT-PCR) was carried out with degenerate oligonucleotides for squalene synthase (SQS), squalene epoxidase  $\beta$ -amyrin synthase and Actin (housekeeping gene). PCR amplification products were of the expected sizes. RT-PCR Products were sequenced to confirm identity and subsequently BLASTn searched across the NCBI database.



**Figure 4.4** RT-PCR using degenerate primers designed from consensus sequences of *M. trunculata*, *P. sativum* and *A. thaliana*. Actin was used as a PCR +ve control. Two sets of oligonucleotide primers were designed per gene. Squalene synthase (SQS F1-R1 and F2-R2), squalene epoxidase (SQE F1-R1 and F2-R2),  $\beta$ -amyrin synthase (BAS F1-R1 and F2-R2) amplified from cDNA produced from Restharrow root. The DNA ladder (1 $\mu$ g per lane) allowed identification of PCR fragment sizes (New England Biolabs; 2 Log DNA ladder), between 10Kb-0.1Kb nucleotides in length, -ve control = non template control.

#### 4.3.8 Design of specific molecular markers from degenerate cDNA amplicons

##### 4.3.8.1 Actin oligonucleotide primers

For use as a PCR positive control, Actin oligonucleotide primers (Table 4.3, ACTIN F1, R1) were designed using a consensus cDNA sequence from *Pisum sativum*, obtained from NCBI nucleotide database, with the Accelrys gene software. The amplicon was excised and purified using Genelute<sup>TM</sup> agarose gel extraction kit and Sanger sequenced to confirm identity (Section 2.2.11).

**Table 4.4** PCR amplified nucleotide sequence for a putative Actin from Restharrow (*O. spinosa*) cDNA using oligonucleotides; Actin F1 and Actin R1, amplification product of 182 nucleic acids.

PCR amplified nucleotide sequence for Actin
TGCTCGAGGGATGCCAAAATGGAGCCTCCAATCCAGACACTATACTTCCTC TCTGGTGGTGCATACAACCTTGAATTTTTCATGCTAGCTTAGGGTGCCAATG CTAGATAATCTCCTTGCTCATCCTATCAGCAATACCAGGGAACATAGTGGT AACCACCACTCAACACAATGTTACCAAG

Table 4.4 shows the nucleic acid sequence from the amplified product Actin F1-R1 (GeneService; Nottingham, UK). The putative ‘Actin’ nucleic acid sequence of 182 nucleotides was BLASTN searched against the NCBI nucleic acid database (figure 4.5). The Actin sequence produced an 89% Maximum identity score, when aligned against U81048.1 (PEAc11), showing a high similarity with the actin nucleic acid sequence in *Pisum sativum*.

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
<a href="#">AC166091.3</a>	Glycine max clone qmw1-11i16, complete sequence	<a href="#">235</a>	235	98%	4e-59	91%
<a href="#">AJ012685.1</a>	Cicer arietinum mRNA for actin, partial	<a href="#">219</a>	219	98%	4e-54	89%
<a href="#">U81048.1</a>	Pisum sativum actin (PEAc11) mRNA, partial cds	<a href="#">219</a>	219	98%	4e-54	89%
<a href="#">U81047.1</a>	Pisum sativum actin (PEAc9) mRNA, complete cds	<a href="#">219</a>	219	98%	4e-54	89%
<a href="#">U81046.1</a>	Pisum sativum actin (PEAc3) mRNA, complete cds	<a href="#">219</a>	219	98%	4e-54	89%
<a href="#">U76192.1</a>	Pisum sativum actin (PEAc11) mRNA, partial cds	<a href="#">219</a>	219	98%	4e-54	89%
<a href="#">U76191.1</a>	Pisum sativum actin (PEAc9) mRNA, complete cds	<a href="#">219</a>	219	98%	4e-54	89%
<a href="#">U76190.1</a>	Pisum sativum actin (PEAc3) mRNA, complete cds	<a href="#">219</a>	219	98%	4e-54	89%
<a href="#">AF143208.1</a>	Vigna radiata actin mRNA, complete cds	<a href="#">213</a>	213	98%	2e-52	88%
<a href="#">U60485.1</a>	Solanum tuberosum actin (Pot66) gene, partial cds	<a href="#">169</a>	169	98%	4e-39	84%
<a href="#">U60484.1</a>	Solanum tuberosum actin (Pot79) gene, partial cds	<a href="#">163</a>	163	98%	2e-37	83%
<a href="#">U60482.1</a>	Solanum lycopersicum actin (Tom52) gene, partial cds	<a href="#">163</a>	163	98%	2e-37	83%

**Figure 4.5** The results from a BLASTN search of the NCBI database, using a putative actin sequence amplified from Restharrow. The top three BLAST hits were from *Glycine max* (AC166091.3), 91% Max identity, *Cicer arietum* (AJ012685.1), 89% Max identity and *Pisum sativum* (U81048.1), 89% Max identity. The NCBI accession number, species/gene information, max score (number of high scoring nucleotide pairs), Query coverage percentage, the expected value and the Max identity (the percentage of high scoring pairs).

#### 4.3.8.2 Squalene synthase Deg2 F2+R2

In order to identify the Restharrow version of squalene cyclase, oligonucleotide primers (Table 4.3, DEG-SQCY-F2 and DEG-SQCY-R2) were designed, using a cDNA nucleic acid sequence from *Pisum sativum*. Sequences were obtained from NCBI nucleotide database, and the Accelrys gene software was used to design the PCR primer pairs. The PCR product was excised and purified using Genelute™ agarose gel extraction kit and Sanger sequenced (Section 2.2.11).

**Table 4.5** PCR amplified nucleotide sequence for a putative squalene synthase from Restharrow (*O. spinosa*) cDNA using oligonucleotides; DEG-SQCY-F2 and DEG-SQCY-R2.

PCR amplified nucleotide sequence for squalene synthase
GAAAGAACCCAAAGTGAAGTATGCAAACGGCATCACGAAGCTCGGTACCAA GCTGTTGAATAACGAGACCGAAGCTTCTAGAAACCTTGTGAAGCATTGAATA ACAGAACGCCCAATGCGGTTCCGGCGGGATCTGCTTCTCGGCTTGCCCTAATC CGCA

Table 4.5 shows the nucleic acid sequence from the amplified product DEG-SQCY-F2 and DEG-SQCY-R2 (GeneService; Nottingham, UK). The putative ‘squalene cyclase’ nucleic acid sequence of 159 nucleic acids, was checked for sequence homology using BLASTN against the NCBI nucleic acid database (figure 4.6). The squalene cyclase nucleic acid sequence produced showed a high level of similarity with the squalene cyclase 1 *Medicago truncatula* (MTR\_4g071520); 90% maximum identity. The BLAST search results suggest that the sequence amplified was likely to be a Restharrow version of squalene cyclase.



Accession	Description	Max score	Total score	Query coverage	E value	Max ident
<a href="#">BT137128.1</a>	<i>Medicago truncatula</i> clone JCVI-FLMT-18F7 unknown mRNA	<a href="#">183</a>	183	89%	3e-43	90%
<a href="#">XM_003606992.1</a>	<i>Medicago truncatula</i> Squalene synthase (MTR_4g071520) mRNA, complete cds	<a href="#">183</a>	183	89%	3e-43	90%
<a href="#">AJ430610.1</a>	<i>Medicago truncatula</i> mRNA for farnesyl-diphosphate farnesyltransferase	<a href="#">183</a>	183	89%	3e-43	90%
<a href="#">AC149293.21</a>	<i>Medicago truncatula</i> clone mth2-78c5, complete sequence	<a href="#">167</a>	167	75%	2e-38	91%
<a href="#">GQ889266.1</a>	<i>Bupleurum chinense</i> clone bcSS1 squalene synthase mRNA, complete cds	<a href="#">165</a>	165	98%	9e-38	84%
<a href="#">AY964186.1</a>	<i>Bupleurum falcatum</i> squalene synthase (SS1) mRNA, complete cds	<a href="#">165</a>	165	98%	9e-38	84%
<a href="#">GQ889267.1</a>	<i>Bupleurum chinense</i> clone bcSS2 squalene synthase mRNA, complete cds	<a href="#">156</a>	156	89%	4e-35	85%
<a href="#">AY787628.1</a>	<i>Centella asiatica</i> squalene synthase (SQS) mRNA, complete cds	<a href="#">154</a>	154	84%	2e-34	85%
<a href="#">XM_003537794.1</a>	PREDICTED: Glycine max squalene synthase-like (LOC100775983), mRNA	<a href="#">150</a>	150	89%	2e-33	84%
<a href="#">GU474427.1</a>	<i>Withania somnifera</i> squalene synthase mRNA, complete cds	<a href="#">150</a>	150	94%	2e-33	83%
<a href="#">GQ889268.1</a>	<i>Bupleurum chinense</i> squalene synthase gene, complete cds, alternative splicing	<a href="#">150</a>	150	83%	2e-33	86%
<a href="#">AK243811.1</a>	Glycine max cDNA, clone: GMFL01-01-F04	<a href="#">150</a>	150	89%	2e-33	84%
<a href="#">AK245602.1</a>	Glycine max cDNA, clone: GMFL01-35-C18	<a href="#">150</a>	150	89%	2e-33	84%
<a href="#">HQ829974.1</a>	<i>Astragalus membranaceus</i> squalene synthase mRNA, complete cds	<a href="#">149</a>	149	93%	7e-33	83%
<a href="#">HQ687217.1</a>	<i>Astragalus membranaceus</i> squalene synthase mRNA, partial cds	<a href="#">149</a>	149	93%	7e-33	83%
<a href="#">GU181386.1</a>	<i>Withania somnifera</i> squalene synthase (SQS) mRNA, complete cds	<a href="#">147</a>	147	94%	2e-32	83%

**Figure 4.6** The results from a BLASTN search of the NCBI database, using a putative squalene synthase sequence amplified from Restharrow (*O. spinosa*). The top two BLAST hits were from *Medicago truncatula* (MTR\_4g071520), 90% maximum identity, *Bupleurum chinense* (bcss1), 84% maximum identity. The NCBI accession number, species/gene information, max score (number of high scoring nucleotide pairs), Query coverage percentage, the expected value and the Max identity (the percentage of high scoring pairs) are listed.

#### 4.3.8.3 Squalene epoxidase oligonucleotide

Utilising genome mining techniques and online genetics tools the isolated nucleic acid sequence from squalene epoxidase in Restharrow was translated into the amino acid sequence. At the amino acid level the squalene epoxidase-like sequence from Restharrow shared a high sequence similarity with squalene epoxidase *Medicago truncatula* (4g092640) and in *A. thaliana* the sequence was most similar to squalene epoxidase 3 (At1g37760).

**Table 4.6** PCR amplified nucleotide sequence for a putative squalene epoxidase from Restharrow (*O. spinosa*) cDNA using oligonucleotides DEG-SQE F2 and DEG-SQE R2.

PCR amplified nucleotide sequence for squalene epoxidase
GNTCGAACTNCTCCAAAGGGATAAGTATAGCACTTAGCTATGTTTTCCAGT CCTTATAGAGAGCATAACCAAACCTACTCTCTTGCAGCATCAATATCAGTCTA CAGCAATCCGTCGAAGCACCCAACTTCAACTTAATCTTGAGATAACCACCAG GTTGTAACAATTACCCACAATTCTATCAGGTTCAATCAAATCTCTTTCAATC ACATGTACTCGCCGTCCATCCTTGCCAAGCGTATAAGCAAGAGCAGA

The Basic Logical Alignment Search Tool, translated protein, (tBLASTx) from NCBI was utilised using a consensus translated amino acid sequence to identify sequence similarities in the genome databases. The alignments show regions of high conservation between the three species *M. truncatula*, *A. thaliana* and *Ononis spinosa*.

**Table 4.7** An Amino acid translation of putative squalene epoxidase (amplified PCR) nucleotide sequence from Restharrow (*O. spinosa*) cDNA using oligonucleotides; DEG-SQE F2 and DEG-SQE R2.

Translated amino acid sequence from putative squalene epoxidase
<u>S</u> A L A Y T L G K D G R R V H V I E R D L N E P D R I V G E L L Q P G G Y L K I K L K L G A S T D C C R L I L <u>M</u> L Q E S S L V <u>M</u> L S I R T G K H S Stop <u>V</u> L Y L S L W

The tBLASTx ref-seq online search tool from NCBI was utilised with a consensus translated amino acid sequence to identify sequence similarities with the *A. thaliana* sequence in genome database. The output in figure 4.7 shows Maximum score, query coverage and a low E\_value indicating a high sequence similarity.

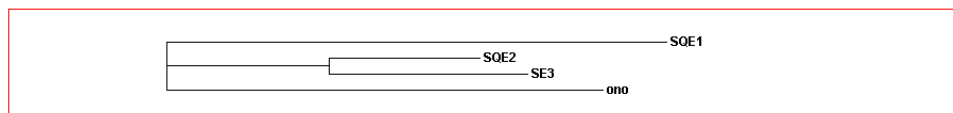
<b>Medicago truncatula MTR 4g092640 tBLASTx 33/35 (94%), 33-16</b>	
Query SQE	T L G K D G R R V H V I E R D L N E P D R I V G E L L Q P G G Y L K I
4g092640	T L G K D G R R V H V I E R D L T E P D R I V G E L L Q P G G Y L K L
	* * *
<b>Arabidopsis thaliana At4g37760 tBLASTx 32/35, (91%), 2e-13</b>	
Query SQE	T L G K D G R R V H V I E R D L N E P D R I V G E L L Q P G G Y L K I
At4g37760	T L G K E G R R V H V I E R D L T E P D R I V G E L L Q P G G Y L K L
	* * *

**Figure 4.7** A tBLASTx amino acid alignment between putative squalene epoxidase from Restharrow (Query SQE) and squalene epoxidase from *Medicago truncatula* (4g092640) and *Arabidopsis thaliana* At3g7760. 94% and 91% homology respectively. \* Indicates amino acid miss match.

Accession	Description	Max score	Total score	Query coverage	E value
<a href="#">NP_564734.1</a>	XF1; squalene monooxygenase [Arabidopsis thaliana]	<a href="#">84.0</a>	84.0	64%	2e-17
<a href="#">NP_179868.1</a>	SQE2 (squalene epoxidase 2); FAD binding / oxidoreductase/ squaler	<a href="#">77.8</a>	77.8	86%	1e-15
<a href="#">NP_568033.1</a>	SQE3 (squalene epoxidase 3); squalene monooxygenase [Arabidopsi	<a href="#">77.0</a>	77.0	61%	2e-15
<a href="#">NP_001031935.1</a>	SQP1; squalene monooxygenase [Arabidopsis thaliana]	<a href="#">69.7</a>	69.7	59%	3e-13
<a href="#">NP_197803.1</a>	SQP1; squalene monooxygenase [Arabidopsis thaliana]	<a href="#">69.7</a>	69.7	59%	3e-13
<a href="#">NP_197804.1</a>	SQE6 (SQUALENE MONOOXYGENASE 6); FAD binding / oxidoreductase	<a href="#">68.2</a>	68.2	59%	1e-12
<a href="#">NP_197802.1</a>	SQP2; FAD binding / oxidoreductase/ squalene monooxygenase [Aral	<a href="#">64.7</a>	64.7	88%	1e-11
<a href="#">NP_680214.1</a>	squalene monooxygenase, putative / squalene epoxidase, putative [	<a href="#">62.8</a>	62.8	61%	4e-11

**Figure 4.8** RT-PCR amplified sequence using primers designed to a squalene epoxidase consensus was BLAST searched against the *A. thaliana* genome database.

Squalene epoxidase nucleic acid sequences (SQE1 (At1g58440), SQE2 (At2g22830) and SQE3 (At4g37760) from *A. thaliana* were translated into amino acid sequences and aligned with a putative squalene epoxidase sequence from *O. spinosa* using ClustalW2 alignment program (<http://EBI.ac.uk>). Squalene epoxidase 3 (At4g37760) showed the highest level of conservation to the putative squalene epoxidase sequence from *O. spinosa*.



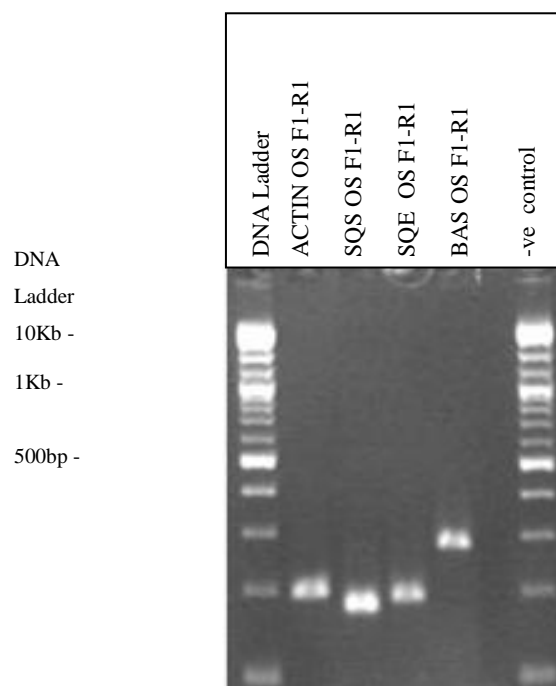
**Figure 4.9** A putative squalene epoxidase (ono) from *Ononis spinosa*, SQE1 (At1g58440), SQE2 (At2g22830) and SQE3 (At4g37760), translated amino acid sequences were aligned using Clustal W2 ([http://www. EBI.ac.uk](http://www.EBI.ac.uk)).

The Squalene epoxidase-like amino acid sequence was isolated from Restharrow and was subsequently subjected to a BLASTP against the *A. thaliana* genome database squalene monooxygenase (known as squalene epoxidase 1 (At1g58440), had the greatest E\_value of 2e-17 and Maximum score of 84. In a study conducted by Matsuda *et al*, (2007), squalene epoxidase; SQE1, (At1g58440), SQE2 (At2g22830) and SQE3 (At4g37760) were cloned and over-expressed in the *Saccharomyces cerevisiae* *Erg1* mutant. The group revealed SQE1 and SQE3 has the ability to produce 2,3-oxidosqualene and 2,3;22,23-dioxidosqualene, whereas SQE2 appeared to produce 2,3-oxidosqualene. Interestingly, single amino acid changes can either increase or decrease the ability of rat ERG1 to metabolize 2,3-oxidosqualene to 2,3;22,23-dioxidosqualene (Abe *et al.*, 2006). Squalene epoxidase 2 (SQE2 (At2g22830), has been functionally characterised in yeast by Rasbery *et al.* (2006). The SQE2 (At2g22830) failed to accumulate detectable amounts of 2,3;22,23-dioxidosqualene (Rasbery *et al.*, 2006). Moreover, certain *Arabidopsis* oxidosqualene cyclases can metabolize, not only 2,3-oxidosqualene, but also 2,3;22,23-dioxidosqualene. (Matsuda *et al.*, 2007).

#### 4.3.8.4 Oxidosqualene cyclases

**Table 4.8** The following species specific primers were designed from the sequenced products amplified using degenerate primers from *O. spinosa* cDNA template. Accelrys Gene software was utilised to develop the primer sequences and theoretical annealing temperatures.

Primer name	Primer sequence	Length	Annealing temp °C
ACTIN-OS-F1	ATGGAGCCTCCAATCCAGAC	20	52.4
ACTIN-OS-R1	GTAACATTGTGTTGAGTGGTGG	22	52.4
SQS-OS-F1	ACCCAAAGTGAAGTATGCAAAC	22	53
SQS-OS-R1	GATTAGGCAAGCCGAGAAG	19	53
SQE-OS-F1	AACCAAACACTCTCTTGCAGC	22	52.3
SQE-OS-R1	TACGCTTGGCAAGGATGGAC	20	52.3
BAS-OS-F1	TGAGGAAGAGAGGGCTCAAG	20	52.3
BAS-OS-R1	GTAGACACAAAGAACCAAGGG	21	52.3
ATSQS-F1	TTGAAAAAGGGTATCAAGAGGC	22	52.5
ATSQS-R1	TCCCAATCTGGTGTCAAAAC	20	52.5
ATSQE1-F1	GAAAGCTGCTTCACTTCCC	19	52.2
ATSQE1-R1	GACCTGAGGATTGCACAAAG	20	52.2
ATBAS-F1	AAACAAGACCCCGAGAGAC	19	52.3
ATBAS-R1	TTCACTGTACTCATGCTCAATC	22	52.3



**Figure 4.10** An agarose gel electrophoresis image showing RT-PCR amplicons from oligonucleotides amplified from *Ononis* genomic sequence. The DNA ladder (1µg per lane) allowed identification of PCR fragment sizes (New England Biolabs; 2 Log DNA ladder), between 10Kb-0.1Kb nucleotides in length, -ve control = non template control. PCR products amplified; housekeeping gene, ACTIN OS F1-R1 ~210bp, SQS OS F1-R1 ~180bp, SQE OS F1-R1 ~190bp, BAS OS F1-R1~290bp.

#### 4.3.8.5 Putative $\beta$ -amyrin synthase

A putative  $\beta$ -amyrin synthase partial sequence was elucidated and supplied by Sean Mayes 2004 (unpublished data) from Restharrow. On the partial sequence, 3' rapid amplification of cDNA ends was carried out to elucidate an almost complete sequence, missing the first 7 amino acid residues on the 5' terminal. The sequence for an extended  $\beta$ -amyrin synthase like- nucleic acid sequence, Metclone 15 can be found in Appendix II.

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
<a href="#">AJ430607.1</a>	Medicago truncatula mRNA for beta-amyrin synthase (bas1 gene)	<a href="#">3169</a>	3169	99%	0.0	92%
<a href="#">AF478453.1</a>	Medicago truncatula beta-amyrin synthase (AMY1) mRNA, complete	<a href="#">3158</a>	3158	99%	0.0	92%
<a href="#">GU072921.1</a>	Glycyrrhiza uralensis isolate 26 beta-amyrin synthase mRNA, complete	<a href="#">2898</a>	2898	99%	0.0	90%
<a href="#">AB037203.1</a>	Glycyrrhiza glabra GgbAS1 mRNA for beta-amyrin synthase, complete	<a href="#">2881</a>	2881	99%	0.0	90%
<a href="#">AB034802.1</a>	Pisum sativum OSCPSY mRNA for beta-amyrin synthase, complete c	<a href="#">2819</a>	2819	99%	0.0	89%
<a href="#">AK339709.1</a>	Lotus japonicus cDNA, clone: LjFL3-062-BD03, HTC	<a href="#">2649</a>	2649	99%	0.0	88%
<a href="#">AF478454.1</a>	Lotus japonicus putative beta-amyrin synthase (AMY1) mRNA, parti	<a href="#">2641</a>	2641	99%	0.0	88%
<a href="#">AB181244.1</a>	Lotus japonicus OSC1 mRNA for beta-amyrin synthase, complete c	<a href="#">2638</a>	2638	99%	0.0	88%
<a href="#">AK245291.1</a>	Glycine max cDNA, clone: GMFL01-26-E16	<a href="#">2567</a>	2567	99%	0.0	87%
<a href="#">AF478455.1</a>	Lotus japonicus multifunctional beta-amyrin synthase (AMY2) mRNA,	<a href="#">2294</a>	2294	98%	0.0	85%
<a href="#">AB034803.2</a>	Pisum sativum OSCPSM mRNA for mixed-amyrin synthase, complete	<a href="#">2060</a>	2060	99%	0.0	83%

**Figure 4.11** A putative  $\beta$ -amyrin synthase-like gene (Sean Mayes (unpublished), 2004), was subjected to an NCBI database BLASTN search. The accession number, species description, maximum score total score, query coverage, E-value (expected value) and maximum identity to the query sequence.

The Restharrow version of a putative  $\beta$ -amyrin synthase nucleic acid sequence was subjected to BLASTN analysis against the NCBI nucleic acid database. The results produced an E-value of 0 for the top 23 BLASTN hits, the top one having a 92% identity to *M. truncatula*  $\beta$ -amyrin synthase (AJ430607.1) as seen in figure 4.11. The Restharrow putative  $\beta$ -amyrin synthase sequence showed homology with *Glycyrrhiza glabra*, *Pisum sativum*, *Lotus japonicas*, and *Glycine max*, indicating a high level of sequence conservation.

An amino acid translation of the  $\beta$ -amyrin synthase-like nucleic acid sequence (unpublished data, Sean Mayes 2004) was carried out (table 4.9) using the ExPASy (<http://web.expasy.org/translate/>) translation tool. The initial 10 amino acids (MWRLKIAEGG) were missing from the 3' terminal and nine coding amino acid residues from the 5'UTR (WALAEYRRRVPLP) in the reference sequence.

**Table 4.9** Putative extended  $\beta$ -amyrin synthase amino acid sequence from Restharrow (from Kloda, 2005). Highlighted in yellow are the QW-motifs, MWCYCR is situated within the active site of the enzyme and SDCTAE has been characterised to have a functional affect in the formation of  $\beta$ -amyrin (Kushiro *et al.*, 2000).

Putative $\beta$ -amyrin synthase amino acid sequence from Restharrow
NDPYLFSTNKFVGRQTWEYDPEAGSEEERAQAEERINFYNNRFEFKPCGDLLWR FQVLRENNFKQTLDGVKIEDGEEITYEKATTTLRRGTHHLAALQTSDGHWPAQIAG PLFFMPPLVFCVYITGHLDVFPQEHKRLRYIYCHQNEDGGWGLHIEGHSTMFC TALNYICMRILGEGPDGGHDNACARARKWIREHGSVTHIPSWGKTWLSILGLFDWC GSNPMPEFWILPSFLPMHPAKMWCYCRRLVYMPMSYLYGKRFVARITPLILELREE LYTQPYEKNVWTKSRHLCAKEDLYYPHPLIQDLIWDSLYIFTEPLLTRWPFNKLVR QKALQVTMKHIHYEDENSRYLTIGCVEKVLCLACWVEDPNGDAFKKHLARVPD YLWMSDGMTMQSFGSQEWDAGFAVQALLATNLVEEIGPALAKGHDFIKKSQVT ENPSGDFKSMHRHISKGSWTFPDQDHGWQVSDCTAEGKCCLLLSLLPPEIVGEK MEPERLYDSVNLNLLSLQSKEGGLSAWEPAGAEWLELLNPTEFFADIVVEHEYVEC TGSAIQALVLFKKLYPGHRKKEIENFIANAVRFLEDTTQTADGSWYGNWGVCFYTG SWFALGGLAAVGKTYGNCAAIRKAVKFLTTQREDGGWGESYLSSPKKIYVPLEG SRSNVVHTAWALMGLIHAGQEERDPTPLHRAAKLLINSQLKEGDWPQEEITGVFM KNCMLHYPMYRDIYPL

The active site of the putative  $\beta$ -amyrin synthase was identified as, MWCYCR (table 4.9). The amino acids SDCTAE was present within the sequence which is thought to play a role in triterpene metabolism (Kushiro *et al.*, 2000). A Protein BLAST of the putative partial  $\beta$ -amyrin synthase-like gene (Kloda, 2005), was carried out to investigate whether the sequence showed greater homology with  $\beta$ -amyrin synthase, lupeol synthase or multi-functional triterpene synthase. The results from the BLASTP (figure 4.12), showed the highly conserved amino acid sequence isolated from *Ononis*, was most similar to  $\beta$ -amyrin synthase from *Medicago truncatula* (Total score 1472), followed by *Pisum sativum* (Total score 1449).

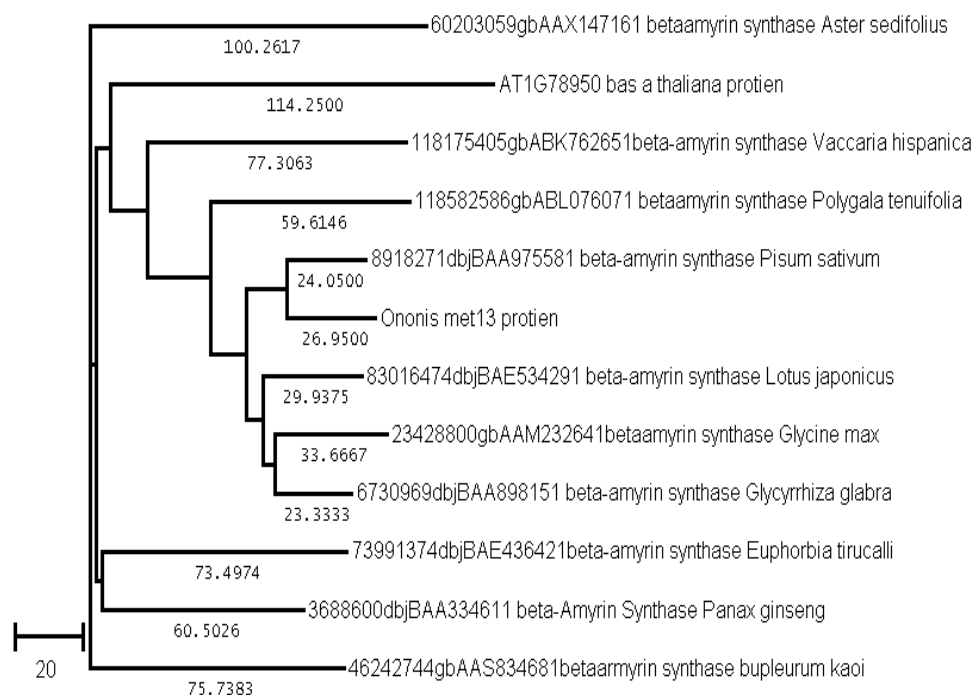


Accession	Description	Max score	Total score	Query coverage	E value
<a href="#">CAD23247.1</a>	beta-amyrin synthase [ <i>Medicago truncatula</i> ]	<a href="#">1472</a>	1472	100%	0.0
<a href="#">AAO33578.1</a>	beta-amyrin synthase [ <i>Medicago truncatula</i> ]	<a href="#">1471</a>	1471	100%	0.0
<a href="#">BAA97558.1</a>	beta-amyrin synthase [ <i>Pisum sativum</i> ]	<a href="#">1449</a>	1449	100%	0.0
<a href="#">ADE88148.1</a>	beta-amyrin synthase [ <i>Glycyrrhiza uralensis</i> ]	<a href="#">1444</a>	1444	100%	0.0
<a href="#">BAA89815.1</a>	beta-amyrin synthase [ <i>Glycyrrhiza glabra</i> ]	<a href="#">1443</a>	1443	100%	0.0
<a href="#">BAE53429.1</a>	beta-amyrin synthase [ <i>Lotus japonicus</i> ]	<a href="#">1428</a>	1428	100%	0.0
<a href="#">AAO33579.1</a>	putative beta-amyrin synthase [ <i>Lotus japonicus</i> ]	<a href="#">1427</a>	1427	99%	0.0
<a href="#">ABL07607.1</a>	beta-amyrin synthase [ <i>Polygala tenuifolia</i> ]	<a href="#">1367</a>	1367	100%	0.0
<a href="#">AAM23264.1</a>	beta-amyrin synthase [ <i>Glycine max</i> ]	<a href="#">1361</a>	1361	100%	0.0
<a href="#">BAB83088.1</a>	beta-amyrin synthase [ <i>Betula platyphylla</i> ]	<a href="#">1337</a>	1337	100%	0.0
<a href="#">AAO33580.1</a>	multifunctional beta-amyrin synthase [ <i>Lotus japonicus</i> ]	<a href="#">1335</a>	1335	100%	0.0
<a href="#">BAE43642.1</a>	beta-amyrin synthase [ <i>Euphorbia tirucalli</i> ]	<a href="#">1303</a>	1303	100%	0.0
<a href="#">ADK12003.1</a>	beta-amyrin synthase [ <i>Aralia elata</i> ]	<a href="#">1303</a>	1303	100%	0.0
<a href="#">XP_002330453.1</a>	predicted protein [ <i>Populus trichocarpa</i> ] >gb EEF08996.1  predicted	<a href="#">1298</a>	1298	100%	0.0
<a href="#">BAF80443.1</a>	beta amyrin synthase [ <i>Bruquiera gymnorhiza</i> ]	<a href="#">1296</a>	1296	100%	0.0
<a href="#">BAA33461.1</a>	beta-Amyrin Synthase [ <i>Panax qinseng</i> ]	<a href="#">1294</a>	1294	100%	0.0
<a href="#">BAA97559.1</a>	mixed-amyrin synthase [ <i>Pisum sativum</i> ]	<a href="#">1286</a>	1286	100%	0.0

**Figure 4.12** NCBI BLASTP (Protein BLAST) of a putative partial  $\beta$ -amyrin synthase-like gene (Kloda, 2005) was carried out to confirm whether the sequence was likely to be a putative  $\beta$ -amyrin synthase. The BLASTP showed the highly conserved amino acid sequence isolated from *Ononis*, was most similar to  $\beta$ -amyrin synthase from *Medicago truncatula* (Total score 1472), followed by *Pisum sativum* (Total score 1449). The figure shows the gene accession, gene description, Maximum score, total score, query coverage and E-value.

#### 4.3.9 Phylogenetic relations between plant oxidosqualene cyclases

Multiple alignments were performed to show similarities between published  $\beta$ -amyrin synthase amino acid sequences, using ClustalW function in Accelrys Gene 2.0 software.  $\beta$ -amyrin synthase is highly conserved, these alignments were then used for nucleic acid alignments between more closely related species to allow degenerate primer design, and produce a neighbour joining phylogenetic tree.



**Figure 4.13** Phylogram showing the relationship between  $\beta$ -amyrin synthase amino acid sequences retrieved from a protein Basic Logical Alignment (BLASTP) from the NCBI database of a putative oxidosqualene cyclase (met13 protein) *Ononis spinosa*.

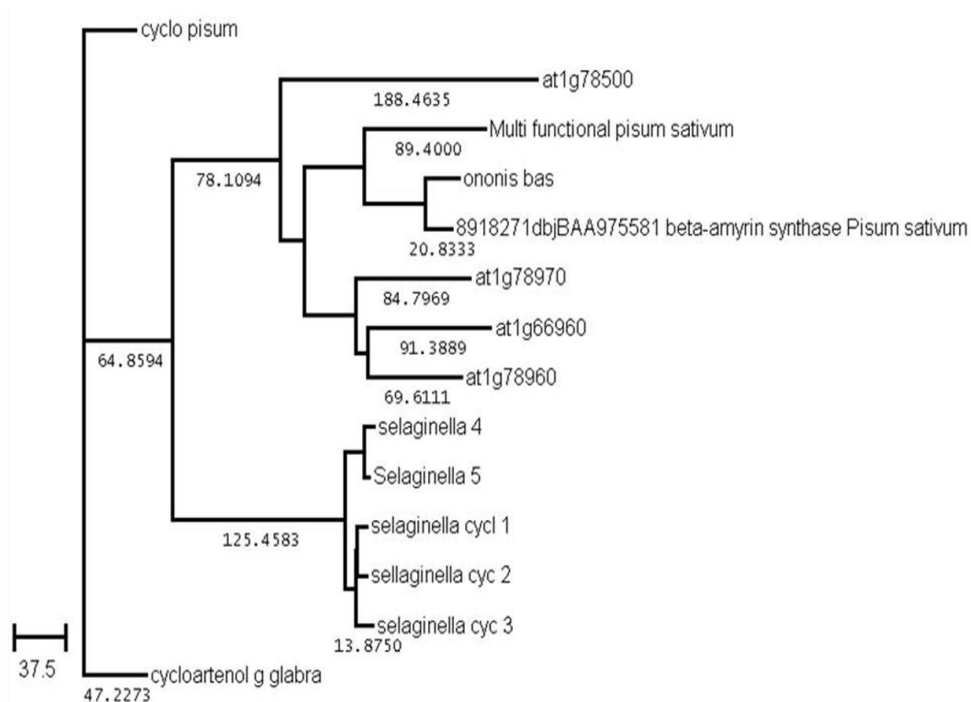
To define the relationships existing between  $\beta$ -amyrin synthases ortholog proteins from angiosperms, a phylogenetic tree analysis based amino acid sequences was calculated according to the Neighbour-Joining algorithm using Accelrys Gene version 2.0 software. *Ononis met13*  $\beta$ -amyrin synthase protein sequence is most similar to  $\beta$ -amyrin synthase *Pisum sativum* (BAA975581). As expected plant species from within the Legume family show most similarity for the putative  $\beta$ -amyrin synthase from Restharrow, *Glycine max*, *Glycyrrhiza glabra*, *Lotus japonicus*.

Amino acid versions of the  $\beta$ -amyrin synthase, lupeol synthase and cycloartenol synthases, from *Pisum sativum* and *A. thaliana* were aligned to compare the active sites and conserved regions. At the amino acid level all oxidosqualene cyclases share a DCTAE at the active site. *Sellaginella*, which has been reported to produce  $\alpha$ -onocerin (Peter Dean, personal

communication) has a L-Serine at residue 499 within the partial sequence. The *Ononis* version of  $\beta$ -amyrin synthase includes an identical amino acid active site sequence to  $\beta$ -amyrin synthase from *Pisum sativum*.

**Table 4.10** A comparison between the active sites within oxidosqualene cyclases between closely related species within the legume family and unrelated *Arabidopsis*. Distantly related to legumes, resurrection plant, *Selaginella*, showed a highly conserved region within the active site.

Amino acid position	1	262	313	496
Clade name	Sites involved in triterpene differentiation			Active site
<i>A. thaliana</i> At1g78970	MWKLK	ILCYSR	YAKED	DCTAE
<i>A. thaliana</i> At1g66960	MWRLK	AFSYTR	CAKED	DCTAE
<i>A. thaliana</i> At1g78960	MWKLK	TLCYTR	CAKED	DCTAE
<i>A. thaliana</i> At1g78500	MWRLK	LWIYFR	CAKED	DCTAE
<i>selaginella</i> 1	MWKLH	MWCHCR	CAKED	DCTAE
<i>selaginella</i> 2	MWKLH	MWCHCR	CAKED	DCTSE
<i>selaginella</i> 3	MWKLH	MWCHCR	CAKED	DCTSE
<i>selaginella</i> 4	MWKLH	MWCHCR	CAKED	DCTAE
<i>selaginella</i> 5	-----	MWCHCR	CAKED	DCTAE
<i>Ononis</i> bas-like	MWRLK	MWCYCR	CAKED	DCTAE
<i>P. sativum</i> BAS	MWRLK	MWCYCR	CAKED	DCTAE
<i>P. sativum</i> MF	MWKLK	MLCYCR	CAKED	DCTAE
<i>P. sativum</i> cycloartenol	MWKLK	MWCHCR	CAKED	DCTAE
<i>G. glabra</i> cycloartenol	MWKLK	MWCHCR	CAKED	DCTAE



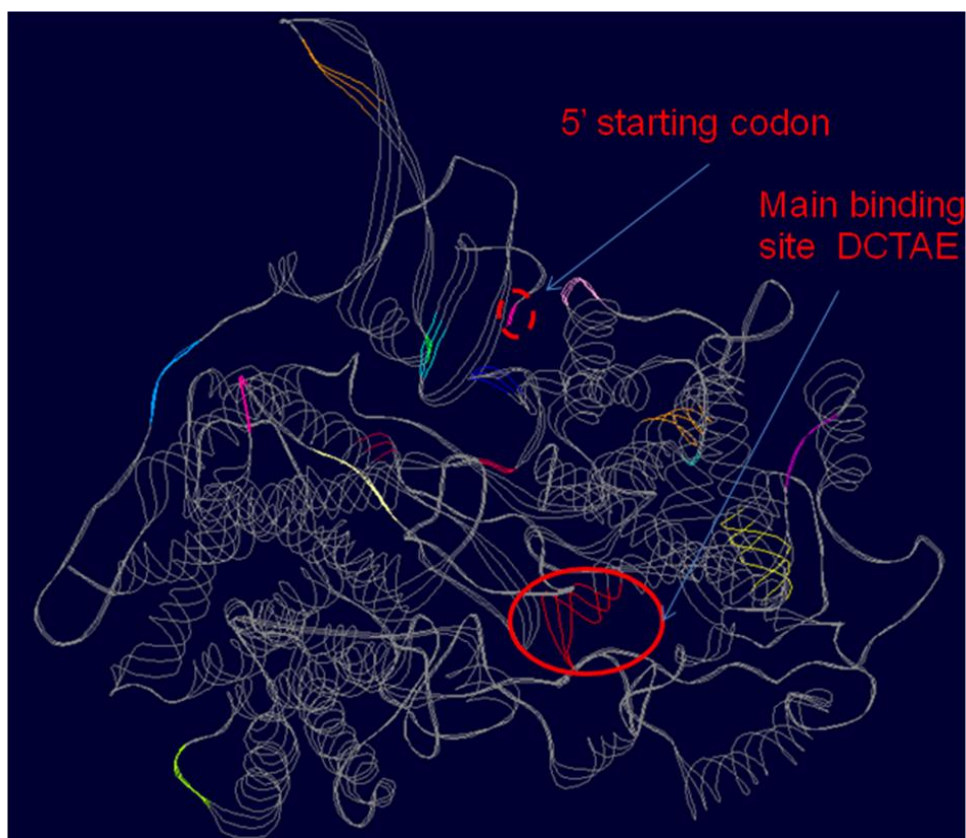
**Figure 4.14** A phylogram showing the relationship between known oxidosqualene cyclases and the putative oxidosqualene cyclase from *Ononis*.  $\beta$ -amyrin synthase (at1g78960), Lupeol synthase (At1g78970 and At1g66960) and Seco-amyrin synthase (At1g78500) cycloartenol synthase (*G. glabra*), cycloartenol synthase,  $\beta$ -amyrin synthase (BAA975581), Multifunctional oxidosqualene cyclase (PSM: BAA97559), from *P. sativum* and cycloartenol synthases and oxidosqualene cyclases from *Sellaginella*.

#### 4.3.10 3D predictive modelling of Oxidosqualene synthases

Amino sequences from three oxidosqualene cyclases were modelled using the 3D predictive modelling Uniprot online modelling tools (<http://www.uniprot.org/help/uniprotkb>). Based on amino acid sequence alignments which showed the highest level of conservation with the *Ononis* putative cyclase (Partial sequence, unpublished data, Mayes 2004),  $\beta$ -amyrin synthase (PSY, BAA97558) and multifunctional triterpene synthase (PSM, BAA97559) from *Pisum sativum* were modelled. The modelling program based the model on that deduced from the human squalene hopene cyclase. The aim was to define the level of conservation at the protein level, and whether the structural similarities between fully

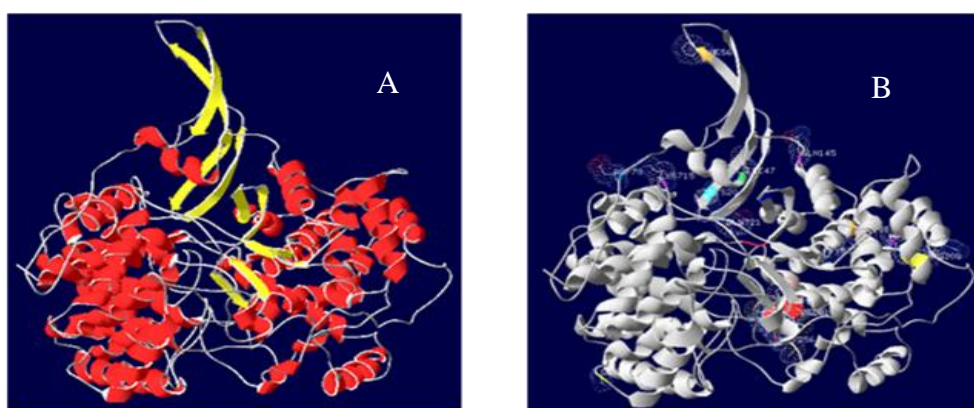
characterised; ( $\beta$ -amyrin synthase; PSY, BAA97558), multifunctional triterpene synthase (PSM, BAA97559) in *P. sativum* and the *Ononis* putative oxidosqualene cyclase (unpublished data, Mayes 2004) were consistent.

The results confirmed that the putative oxidosqualene cyclase from *Ononis* was similar and single amino acid differences within the protein sequence within *Ononis*, compared to other functionally characterised sequences, had little effect on the structural conformation. The results also allowed important regions within the amino acid sequence to be mapped such as the DCTAE (*Ononis*) binding site, the 5' start codon,  $\alpha$ -helix and  $\beta$ -pleated sheets (figure 4.18). The PdB viewer allowed simultaneous modelling between several amino acid sequences, to allow a direct comparison of predicted structural conformation (figure 4.18). There was little difference between the secondary and tertiary structures ( $\alpha$ -helixes and anti-parallel  $\beta$ -strands). The main differences occurred between the BAA97559 sequence in *P. sativum* and oxidosqualene cyclase in *Ononis*, as indicated in figure 4.18. Remarkably the oxidosqualene cyclase 3D protein model from *Ononis* was almost identical to  $\beta$ -amyrin synthase; PSY, BAA97558). There was no obvious structural differences or differences within the active site; DCTAE region.

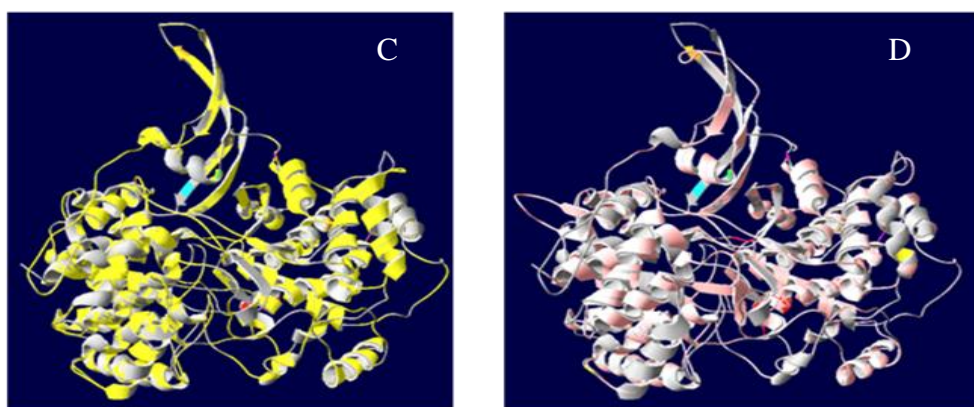


**Figure 4.15** 3D predictive modelling of an *Ononis* putative oxidosqualene cyclase (unpublished data, Mayes 2004). Output from Swiss Prot-automatic output viewer. PdB file, manipulated using Swiss model pdb viewer.

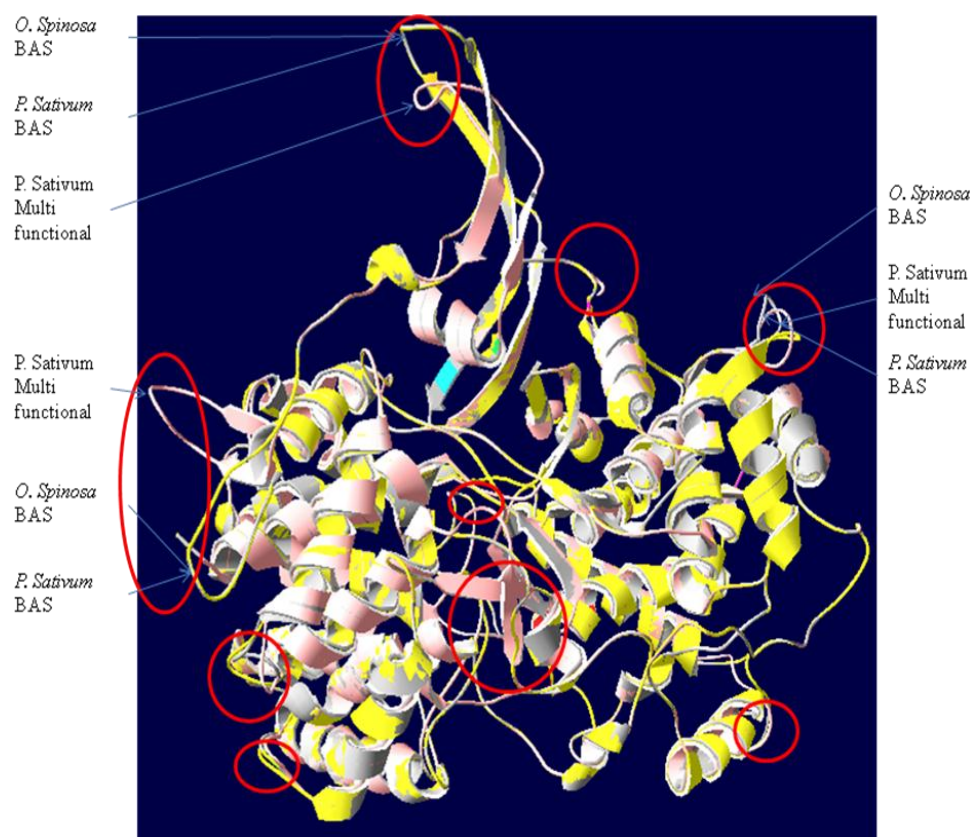
There were 13 amino acid differences between  $\beta$ -amyrin synthase-like protein sequence oxidosqualene cyclases across 15 genes ( $\beta$ -amyrin synthases and multi-functional triterpene synthases). The regions highlighted (coloured regions), did not appear to affect the physical structure of the active site. Red coloured region represent the binding sites.



**Figure 4.16** Secondary protein structure of (A)  $\beta$ -amyrin synthase (*P. sativum*; BAA97558). Red represents alpha helix structures encoded by QW motifs and yellow regions represent  $\beta$ -pleated sheets. (B) represents putative oxidosqualene cyclase (*O. spinosa*). Differences between the amino acid sequences were highlighted. Output PdB files: presented in a Swiss model PdB viewer (<http://www.uniprot.org/help/uniprotkb>).



**Figure 4.17** Secondary protein structure of (C)  $\beta$ -amyrin synthase (*P. sativum*, BAA97558) overlaid with a putative oxidosqualene cyclase (*O. spinosa*). A (D) multifunctional oxidosqualene cyclase (*P. sativum*, BAA97559) was directly modelled against a putative oxidosqualene cyclase (*O. spinosa*). The sites of variation are marked, indicating slight differences in amino acid sequence.



**Figure 4.18** Structural predictive 3D protein models of a putative oxidosqualene cyclase (*O. spinosa*) a  $\beta$ -amyryn synthase (*P. sativum*; PSY, BAA97558), and a multifunctional oxidosqualene cyclase *P. sativum*; PSM, BAA97559) modelled using Expasy SWISSprot modelling software. Major structural differences are indicated between the three protein models.



#### 4.4 DISCUSSION

The use of bioinformatics for mining plant metabolomes and plant genomes may offer several key starting points for investigating  $\alpha$ -onocerin production in Restharrow. There are several stages within the Triterpene pathway which could be modified in Restharrow to allow  $\alpha$ -onocerin production.

The Aracy tool from NASC (<http://arabidopsis.info/>) gave genome-based information on the genes involved in triterpene biosynthesis in several plant species which have been more fully characterised. In *A. thaliana* and *O. sativa*, there are two copies of squalene cyclase, whereas in closer relatives to Restharrow, *Panax ginseng* and *Medicago truncatula*, there is only a single copy. *M. truncatula* belongs to the legume family and of the for plant species, it is most closely related to Restharrow.

Aracyc results showed there are six squalene epoxidases in *A. thaliana* (Brassicales). Three squalene epoxidases have been characterised by functional expression studies (Rasbery *et al.*, 2007); At1g58440, At2g22830 At5g24160, all of which are located on separate chromosomes. At5g24140, At5g24150 and At5g24160 all occur as a gene cluster on chromosome 5 which suggests gene duplication may have occurred. *Panax notoginseng* (Apiales) has a single squalene epoxidase (Pn-SE), whereas in *Solanum Lycopersicum* (Solanales) four squalene epoxidases (SI-SGN-U315271, SI-SGN-U315272, SI-SGN-U317223, SI-SGN-U335064), have been reported. Two squalene epoxidases were present in *M. truncatula*, (AJ430609 and AJ430608).

In *Glycyrrhiza glabra* (Fabales; Gg-bAS1-BAA898) *Medicago truncatula* (Fabales; Mt-bas1-CAD23247), and *Avena strigosa* (Poales; ASBAS1-CAC84558), there is a single  $\beta$ -amyrin synthase gene. One of the most closely related species to Restharrow is pea (*Pisum sativum*) which has a single copy of  $\beta$ -amyrin synthase as discussed in chapter 1 (PSY, BAA97558), although also has a mixed oxidosqualene synthase (PSM BAA97559) which has been reported to cyclase  $\beta$ -amyrin (Morita *et al.*, 2000). The fact that there is a single  $\beta$ -amyrin synthase present in *P. sativum*

may have important implications when considering cyclisation mechanisms which may occur in Restharrow. *Panax ginseng* (Apiales) has two copies of  $\beta$ -amyrin synthase (Pg-OSCPNY1-BAA33722, BAA33461). In the case of *Arabidopsis thaliana* there are several genes closely related to one another which show very high levels of similarity and produce  $\beta$ -amyrin as a product. The primary gene responsible for cyclising 2,3-oxidosqualene into  $\beta$ -amyrin is  $\beta$ -amyrin synthase At1g78950, although lupeol synthases At1g78960, At1g78970, have been reported to cyclise the monoepoxide to  $\beta$ -amyrin. At1g66960 a lupeol synthase, and an At1g78500 have been reported to cyclise the monoepoxide to  $\alpha$ -amyrin.

In *A. thaliana* within the triterpene pathway there are many examples where gene duplication may have occurred. Gene duplication results in an additional copy of a gene situated on either the same chromosome or on a separate chromosome. The resulting gene duplication allows the additional gene to become free from selective pressure (if the original gene function is an essential one) to mutate without deleterious consequences within the organism. Free from deleterious consequences the duplicated gene may mutate to have a novel function, potentially increasing the fitness of the organism under selection. Gene duplications may occur via a number of mechanisms; an error in homologous recombination, a retrotransposition event or a duplication of an entire chromosome.

There are several examples of where gene duplication may have occurred within the triterpene pathway in *A. thaliana*. By looking at chromosomal locations and linking them to gene function it is possible to tell where gene duplication may have occurred in a pathway context. The two genes closely linked occur on chromosome 4 have been characterised as squalene synthases; At4g34640 and At4g34650 with a single function. Six squalene epoxidases also known as squalene monooxygenases have evolved in *A. thaliana* with three have been characterised by functional expression studies in yeast; At1g58440, At2g22830 At5g24160, all of which occur on separate chromosomes. At5g24140, At5g24150 and At5g24160 all occur

closely together on chromosome 5, suggestive of a duplication event (Phillips *et al.*, 2006).

By carrying out genome mining techniques it has highlighted the fact that there may be some redundancy in the triterpenoid pathway and possibly differences in gene copy number between *A. thaliana* and legumes: specifically *A. thaliana* and *M. truncatula*. Although the genome has been fully sequenced and there are several publications concerning the biosynthetic mechanism for triterpene biosynthesis in legumes such as *M. truncatula* and *P. sativum* there are still relatively few genomes fully characterised. For this reason from the results of the genome mining experiments, it was considered that *A. thaliana* is a suitable model to be used for creating a gene atlas of genes involved in triterpene synthesis. The apparent gene duplication present in Arabidopsis, compared to the known legume species is an important consideration, but the availability of resources for comparative analysis in *A. thaliana* makes it the most comprehensive model available for this pathway.

The genetic information available from the model crop *A. thaliana* proved to be a valuable tool in developing gene targets to follow triterpene biosynthesis in Restharrow. The initial squalene synthase, squalene epoxidase,  $\beta$ -amyrin synthase and other oxidosqualene cyclase sequences, were searched using a BLASTN search initially at the nucleic acid level. Nucleic acid sequences for genes with a close match to the *A. thaliana* gene versions and legumes were translated into amino acid sequences. From these multiple alignment translated amino acid consensus sequence, genetic markers specific for Restharrow were developed. Degenerate primers successfully produced PCR products of the expected sizes as a result of RT-PCR. The sequence results allowed the development of molecular markers to specifically target putative *O. spinosa* squalene synthase, squalene epoxidase and  $\beta$ -amyrin synthase. These molecular tools would provide a starting point for transcriptional studies within Restharrow and functional expression studies in either plants or yeast.

From the results of amino acid alignments it was deduced that there was a higher level of conservation between the putative squalene epoxide (Restharrow) and squalene epoxidase 3 (At4g37760). Squalene epoxidase 3 (At4g37760) and Squalene epoxidase 1, (At1g58440) in *A. thaliana* are the primary enzymes involved in cyclising 2,3-oxidosqualene to 2,3;22,23-dioxidosqualene. It is feasible for a squalene epoxide to play a similar role in Restharrow. This identified an area in the current study to be investigated further using transcriptional profile.

From analysing the data within plant metabolomic data bases, a single copy of the  $\beta$ -amyrin synthase gene has been identified in *Glycyrrhiza glabra*, *Panax ginseng* and *Medicago truncatula*. *Medicago truncatula* is a closely related species to Restharrows both being classified within the *Leguminosae* family. *Pisum sativum* also has a single copy of beta-amyrin synthase, although a mixed amyrin synthases is also present. This suggests that there may be a single copy of  $\beta$ -amyrin synthase in Restharrow and possibly a secondary multifunctional triterpene synthase. This inference may prove to be significant and suggests a single copy of a  $\beta$ -amyrin like triterpene synthase may be present in Restharrow. As such the evidence presents a strong case that the production of  $\alpha$ -onocerin may be substrate dependant.

The putative  $\beta$ -amyrin synthase (partial), sequence elucidated and supplied by Sean Mayes 2004 (unpublished data) from Restharrow was extended towards the terminal 5' UTR to include the **WALAEYRRRVPLP**. The 3' UTR is yet to be elucidated. At the nucleic acid level and the amino acid levels the putative  $\beta$ -amyrin synthase (partial), was most similar to the *P. sativum* version of the  $\beta$ -amyrin synthase (BAA975581), rather than the multifunctional triterpene synthase (BAA97559). This might suggest that in Restharrow this particular gene encodes a  $\beta$ -amyrin synthase rather than acting as a multifunctional oxidosqualene cyclase. The putative  $\beta$ -amyrin synthase from Restharrow can be considered as a valid candidate for future functional characterisation studies. Without full functional expression studies of the putative  $\beta$ -amyrin synthase (Restharrow) it is impossible to

predict the products directly produced by the enzyme. It is notable however that the sequence was highly conserved at the amino acid level to BAA975581 at 93% similarity compared to a 79% sequence similarity to BAA97559. What this work highlights is a potential candidate gene that may be involved in the triterpenoid biosynthetic pathway in *Restharrow*.

The putative  $\beta$ -amyrin synthase (partial) amino acid sequence from *Restharrow* was directly compared to sequences from active sites from a range of oxidosqualene cyclases: *A. thaliana* at1g78970 (lupeol synthase) at1g66960 (multifunctional seco amyrin synthase), At1g78960 ( $\beta$ -amyrin synthase), At1g78500 (multifunctional synthase); *P. sativum*- BAA975581 ( $\beta$ -amyrin synthase), BAA97559 (Multi-functional) *P. sativum* (cycloartenol synthase), *G. glabra* (cycloartenol synthase) and three putative cycloartenol synthases from *Selaginella* which is reported to accumulate  $\alpha$ -onocerin (Peter Dean Pers. comm.). All oxidosqualene cyclases had the same amino acid sequences in the region which has been reported as the active site (Abe *et al.*, 1993; Abe and Prestwich, 1995); 'DCTAE', with of exception of two of the *Selaginella* cycloartenol synthases which had a sequence of DCTSE (table 4.6). SDCTAE is so conserved in  $\beta$ -amyrin sequences that Corey (1997) argues SDCTAE could be involved in the protonation of the oxirane which initiates cyclization. The protein sequences include a number of hydrophilic and hydrophobic domains, although the hydrophobic runs are often too short to constitute membrane-spanning domains. The incomplete sequence shares the presence of 5 QxxxxxW motifs concentrated at the 5' and 3' ends of the sequence. These motifs may be important in the stabilization of carbocationic intermediates of OSC cyclization (Porrilla, 1994). Another protein region which has been reported to play important roles in the formation of  $\beta$ -amyrin and other triterpenoids (Kushiro *et al.* 2000), involves the 'MWCYCR' sequence. The tryptophan residue in the MWCYCR motif of the pea  $\beta$ -amyrin synthase has been identified by site-directed mutagenesis as being important for  $\beta$ -amyrin formation and is conserved amongst all other published  $\beta$ -amyrin synthase sequences (Kushiro *et al.*, 2000; Harampidis *et al.*, 2002).

Based on the same amino acid sequences, remarkably the oxidosqualene cyclase 3D protein model from *Ononis* was almost identical to  $\beta$ -amyirin synthase; PSY, BAA97558). There were no obvious structural differences within the active site; DCTAE or other important regions. The 3D predictive modelling provided further evidence to suggest the partial putative  $\beta$ -amyirin synthase encodes a functional  $\beta$ -amyirin synthase in Restharrow.

## CHAPTER 5. OCCURRENCE OF $\alpha$ -ONOCERIN AND CYCLIC PRECURSORS IN RESTHARROW

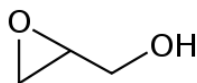
### 5.1 INTRODUCTION

#### 5.1.1 General introduction

Plants produce a large number of heterocyclic triterpene derived products as part of their normal growth and development. The elucidation of biological function and biochemical cyclisation mechanisms of such metabolites using genetics and biochemical approaches has received considerable attention (Corey *et al.*, 1992; Morita *et al.*, 2000; Haralampidis *et al.*, 2001; Husselstein-Muller *et al.*, 2001; Basyuni *et al.*, 2006; Qi *et al.*, 2004). Despite this research effort, the biosynthetic mechanisms and ecological function of such a structurally diverse group of naturally occurring plant secondary metabolites (estimated 20,000 different structures), are still poorly understood (Phillips *et al.*, 2006). The triterpene  $\alpha$ -onocerin for example, has been reported as having medicinal (Orhan *et al.*, 2005; 2006; Rollenger 2005), antifungal and antibacterial properties (Mahasneh and El-Oqlah, 1999) however, the cyclisation mechanism is still poorly understood.

#### 5.1.2 Triterpene biosynthesis in plants: Squalene cyclisation

Squalene cyclases have evolved to protonate the oleofinic functional group which is less basic than the hydroxyl (OH) group. One example of this is a tetrahymenol synthase, which normally cyclises squalene into tetrahymenol, but can also convert 2,3-oxidosqualene to 3,21-gammaceranediol (Bouvier *et al.*, 1980). Essentially the tetrahymenol synthase can cyclise both a squalene derived product and an 2,3-oxidosqualene product. Squalene derived triterpenes such as tetrahymenol and hopanol are found in prokaryotes and ciliates (Holtvoeth *et al.*, 2010). X-ray crystallography studies suggest the cyclisation of squalene generates dammarenyl cations only (figure 5.2). Squalene is a symmetrical molecule and therefore can be cyclised from both termini (Haralampidis *et al.*, 2002).

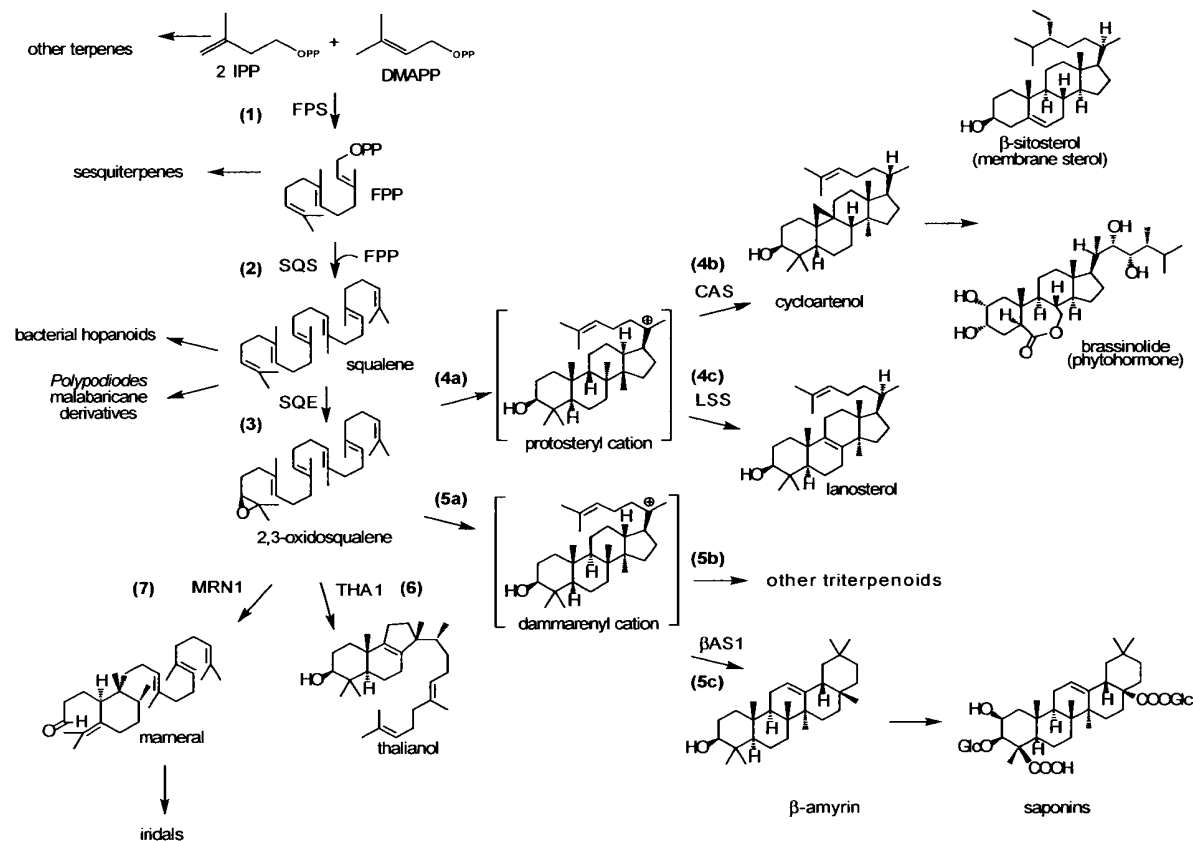


**Figure 5.1** The triangular carbon-carbon-oxygen based terminal formed by the epoxide group affords a reactive intermediate when attached to squalene. This aids the cyclisation of squalene epoxide from the epoxy terminal.

### 5.1.3 Cyclisation of 2,3-oxidosqualene

Originally proposed by the biogenetic isoprene rule (Ruzicka *et al.*, 1953; Abe *et al.*, 1993), cyclisation and deprotonation reactions allow differential products (as shown in figure 5.2) to be produced. 2,3-oxidosqualene cyclase (OSC) generates numerous tetra-terpene and tri-terpene alcohols because tetracyclic cations are often subjected to hydroxylation or deprotonation prior to neutralisation. Tetracyclisation events produce two different types of intermediate precursors that differ only in B ring conformation. The protosteryl cation has a chair-boat-chair and the dammarenyl cation has a chair-chair-chair conformation. Phytosterol biosynthesis is initiated by enzymatic cyclisation of 2,3-oxidosqualene, prior to a ‘chair-boat-chair’ conformation to yield a C<sub>20</sub> protosteryl cation, which is subsequently converted into cycloartenol or lanosterol by cycloartenol synthase or lanosterol synthase, respectively. (Abe *et al.*, 1993; Shibuya *et al.*, 2007).





**Figure 5.2** A representation of triterpene biosynthesis as it occurs in *A. thaliana*. Products derived from the protosteryl cation: cycloartenol, lanosterol, β-sitosterol and brassinolide. Triterpene products derived from dammarenyl cation; β-amyrin and saponins.

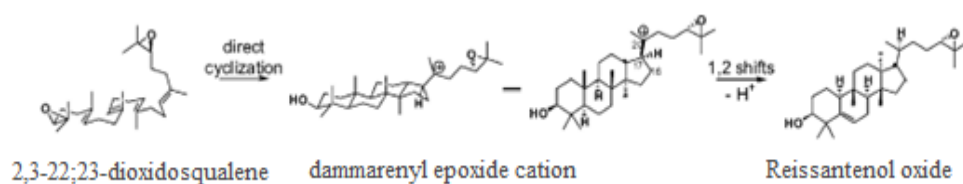
The dammarenyl cation may be converted to dammarane like triterpenes by 2,3-oxidosqualene cyclase - dammarenediol synthase (van Tamelen *et al.*, 1966) or may undergo further cyclisation by enzymes such as lupeol synthase and  $\beta$ -amyrin synthase (Yoder, 2005). A more diverse range of triterpenoid secondary metabolites can be found when 2,3-oxidosqualene cyclisation is the primary route towards triterpene cyclisation; primarily in angiosperm plants (Suzuki, 2002). Thus angiosperms such as Restharrow are good organisms in which to conduct triterpene distribution, accumulation and biosynthesis studies

#### **5.1.5 Natural products derivitised from 2,3;22,23-dioxidosqualene**

Two modes of cyclisation from a 2,3;22,23-dioxidosqualene have been reported in the literature. The first mechanism involves the cyclisation of an entire 2,3;22,23-oxidosqualene product, followed by further post cyclic modifications by cytochrome P450 enzymes. The second mechanism involves a partial cyclisation retaining the 18,19 olefin unsaturated carbon chain, with a second cyclisation initiated from the distal 22,23-epoxide, resulting in the formation of epoxyalcohols. Some natural products undergo a single electrophilic attack, with a carbocation generated by the initial cyclisation event producing five and six-membered oxirane structures.

Attention has been focussed on investigating 2,3;22,23-dioxidosqualene cyclisation into a 24,25-epoxylanostan-3-ol product (Field and Holmlund, 1977). Within yeast (Field and Holmlund, 1977), and animals (Spencer, 1994; Gardener *et al.*, 2001), under normal growth conditions, squalene is cyclised into cycloartenol by lanosterol synthase. In conditions where the expression level of lanosterol synthase is limited, the distal terminus of the 2,3-oxidosqualene has been shown to re-enter the squalene epoxidase enzyme complex to add a hydroxyl (OH) group to the 23, 22 carbon molecule (epoxy-lanosterol). According to Spencer (1994), the epoxy-lanosterol accumulation acts as a negative feedback signal in which reissantenol regulates sterol levels in mammals and yeast.

Remarkably, a similar mechanism has evolved in plants where a cycloartenol synthase cyclases 24,25-(S)-epoxycycloartan-3-ol (De Pascual Teresa *et al.*, 1987; Della Greca *et al.*, 1994), which is the epoxy derivative of the corresponding plant sterol precursor cycloartenol. Reissantenol oxide has been isolated from the root bark of *R. indica* (Gamlath *et al.*, 1989) is the 24,25-(S)-epoxide of a rearranged dammarane euferol. 2,3-22;2,3-dioxidosqualene has been shown to be further cyclised into oxysterols such as 24,25-epoxycholesterol (Nelson *et al.*, 1981).



**Figure 5.3** A representation of the conversion of 2,3;22,23 - dioxidosqualene to reissantenol oxide by a direct cyclisation event.

### 5.1.5 Occurrence of triterpene diols

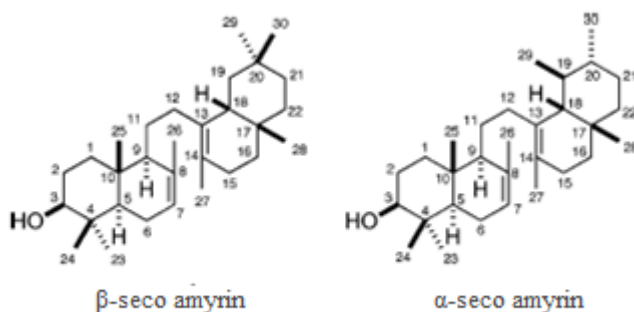
To date there is little evidence to suggest triterpene diols, which are products of the triterpene pathway, occur regularly in nature. The characterisation of *A. thaliana* LUP1 (At1g78970), discovered the accumulation of multiple triterpene products when expressed in a yeast lanosterol synthase mutant SMY8. Among them were;  $\beta$ -amyrin, taraxasterol and germanicol (a 3,20-dihydroxylupane), (Segura *et al.*, 2000). The study showed direct evidence of a 3,20-dihydroxylupane, and suggested that it may occur naturally in *A. thaliana*. The cyclisation mechanism of 3,20-dihydroxylupane is unclear and yet to be elucidated, although it has been reported in six plant species from the Asteridae and Rosidae families.

A pentacyclic triterpene diol; 13-Serratene-3b,21a-diol (Inubushi *et al.*, 1964), contains an unusual seven membered C-ring and occurs in *Lycopodium spp.* (Orito *et al.*, 1972). 13-Serratene-3b,21a-diol may be cyclised by a route similar to that established for  $\alpha$ -onocerin during the cyclisation cascade from a pre-onocerin intermediate. Spruce trees contain natural 13-serratene-3,21-diols, these compounds are likely to be derivatives

of 13-serratene-3b,21a-diol. A similar process is believed to generate diverse conifer ahydroxytriterpenoids (Inubushi *et al.*, 1962; Xu *et al.*, 2004).

#### 5.1.6 Post cyclic modifications of triterpenes

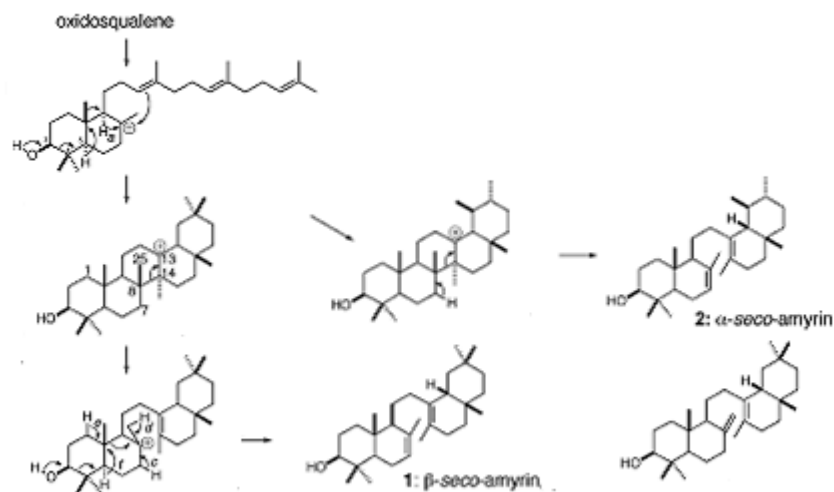
In *A. thaliana* 2,3-oxidosqualene cyclase (At1g78500) has been shown to produce a sub-set of triterpenes which are 6-6-5-6-6 structural conformations; ‘seco-triterpenes’, which have undergone C ring cleavage. The same 2,3-oxidosqualene cyclase enzyme has been shown to produce;  $\beta$ -amyrin,  $\alpha$ -amyrin, multiflourenol, baurenol and  $\beta$ -seco amyrin and  $\alpha$ -seco amyrin. One explanation for the biosynthesis of seco-amyrin is the re-entering of substrate into the 2,3-oxidosqualene synthase active site to undergo C ring cleavage (Shibuya, 2007).



**Figure 5.4** Stereo-chemical structure of  $\beta$ -seco amyrin and  $\alpha$ -seco amyrin. The numbers indicate the positions of Carbon, HO indicates a Hydroxyl group.

Seco-triterpenes, are increasingly being discovered to be present in plants and are being characterised using functional expression studies coupled with mass spectrometry. Recently, Shibuya *et al.*, (2007) reported a 2,3-oxidosqualene cyclase from *A. thaliana*, identifying it as marnerial synthase (At5g42600), yielding the A-ring-*seco*-monocyclic triterpene (figure 5.4). In the case of the At1g78500-catalyzed reaction, a carbocation returns to C13 of the C-ring in a series of anti-parallel hydride and/or methyl group shifts from pentacyclic carbocation intermediates. A cleavage of the carbon-carbon bond between C8 and C14 to form a C13-C14 double bond also occurs to form C-ring-*seco*-amyrrins (Shibuya *et al.*, 2007). Further

interesting *seco*-triterpene structures include those reported from *Camellia* species i.e., camelliol.



**Figure 5.5** The cyclisation of alpha and beta seco-amyrins from a 2,3-oxidosqualene derivative.

#### 5.1.7 Proposed mechanistic pathways for $\alpha$ -onocerin biosynthesis

From the three hypotheses listed in Chapter 1.14 the following observations were expected from metabolomic studies as listed in table 5.1. If 2,3-oxidosqualene cyclase was the primary precursor 2,3-oxidosqualene, pre-onocerin or seco-amyrin or an alternative oxidosqualene diol intermediated would be expected. The results would show the presence of 2,3;22,23-dioxidosqualene, where  $\alpha$ -onocerin occurs, if a 2,3;22,23-dioxidosqualene was the primary candidate for cyclisation. Alternative intermediate metabolites may be seen if the cyclisation event is part of a different pathway.

**Table 5.1** Possible outcomes for three hypotheses for  $\alpha$ -onocerin cyclisation within Restharrow. Presence or absence of triterpenes and their precursors from within plant extracts.

<b>Expected compounds from extracts</b>	<b>Hypothesis 1</b> (Initial precursor 2,3-oxidosqualene)	<b>Hypothesis 2</b> (Initial precursor 2,3;22,23-dioxidosqualene)	<b>Hypothesis 3</b> (Different precursor)
2,3-oxidosqualene	Yes	Yes	Yes
2,3;22,23-dioxidosqualene	No	Yes	No
Pre-onocerin	Yes	No	No
seco-amyrin	Yes	No	No
$\beta$ -amyrin-diol/taraxasterol-2,3,21-diol	Yes	No	No
Increased accumulation of $\beta$ -amyrin or $\alpha$ -onocerin	Yes	Yes	Yes

### 5.1.8 Summary

Little is known about how  $\alpha$ -onocerin, its precursors or related triterpenoid product accumulation changes between species or throughout development within Restharrow. By investigating the biochemical content of Restharrow and other species which either do or do not accumulate  $\alpha$ -onocerin as a secondary metabolite, using a comparative approach, it may provide an insight into the triterpene biosynthetic pathways in plants.

In this chapter, analytical chemistry techniques such as extraction, derivitization, method development, GC-MS, identification and quantification of spatially and temporally sampled Restharrow were used as a way to investigate the cyclisation of such biochemical molecules. The study focussed on measuring the accumulation of triterpene precursors 2,3-

oxidosqualene and 22,23:2,3-dioxidosqualene and triterpenoids;  $\alpha$ -onocerin and  $\beta$ -amyrin.

## **5.2 RESULTS: DERIVITISATION**

### **5.2.1 Identification of internal standards from plant extracts**

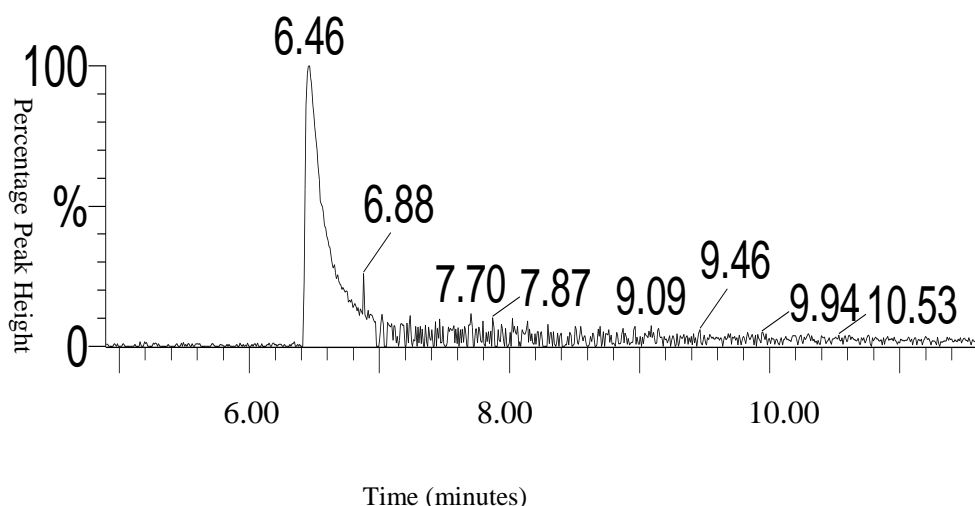
Internal standards were derivitised for use as controls within all experiments. Adding a known concentration to each of the experimental samples corrected for between experiments, extraction efficiency and sample variation. Normalisation techniques could subsequently be used for quantification purposes. Methods were developed for detecting the three alkanes used as internal standard controls; hexacosane ( $C_{26}$ ), octacosane ( $C_{28}$ ) and triacontane ( $C_{30}$ ) suspended in dichloromethane. Prior to injection three consecutive blank injections of dichloromethane were added to clear the column of unwanted residue and to test for solvent contamination. A manual liquid injection of 1  $\mu$ l of sample into an Electron Impact Ionisation Mass Spectrometer (EIMS), was used in tandem with gas chromatography to produce a definitive mass spectrum for each of the standards. The Mass spectra were directly compared with standards listed on the National Institute of Standards (NIST) mass spectral library.

The retention time was used as a presence/absence indicator and the mass spectrum was used to confirm the identity of the analyte of interest. Figure 5.6 and 5.7 showed that alkanes;  $C_{26}H_{54}$  (hexacosane), retention time 5.46 minutes,  $C_{28}H_{58}$  alkane (octacosane), retention time 7.59 minutes,  $C_{30}H_{62}$  (triacontane), retention time 9.04 minutes, could be used as internal standards. The analyte's molecular weights were initially compared at 366.71 (hexacosane), 394.77 (octacosane) and 422.83 (triacontane). This comparison allowed a prediction that they would occur along the chromatogram in the same order according to the molecular weight. triacontane was selected to be used as an internal standard as it occurred at an appropriate retention time so as not to interfere with other analytes within the fraction.

**Table 5.2** The physical properties of three long chain carbon molecules hexacosane, octacosane, triacontane belonging to the group; alkanes, used as internal standards for quantification of triterpenes. The table outlines the boiling point and melting point (Degrees Celsius), the density (grams per centimetre cubed) in relation to water and the molecular weight.

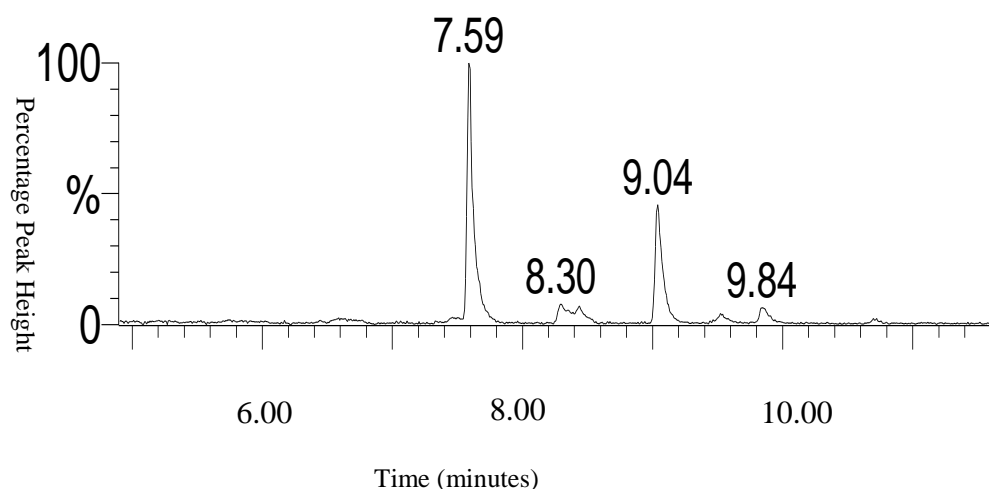
Property	Hexacosane	Octacosane	Triacontane
Boiling point	420 °C	440 °C	450 °C
Melting point °C	55-58	61-64	66
Density (g/cm <sup>-3</sup> ) H <sub>2</sub> O=1	0.8	0.777	0.775
Molecular weight g/mol	366.71	394.77	422.83

The physical properties of internal standards were used to identify temperature, extraction procedures and identify, an understanding of where they may occur on the Chromatogram. The information in table 5.2 facilitated the identification of analytical controls by the identifying the molecular ion. The molecular ion corresponds to the molecular weight, aiding identification within the mass spectrum.



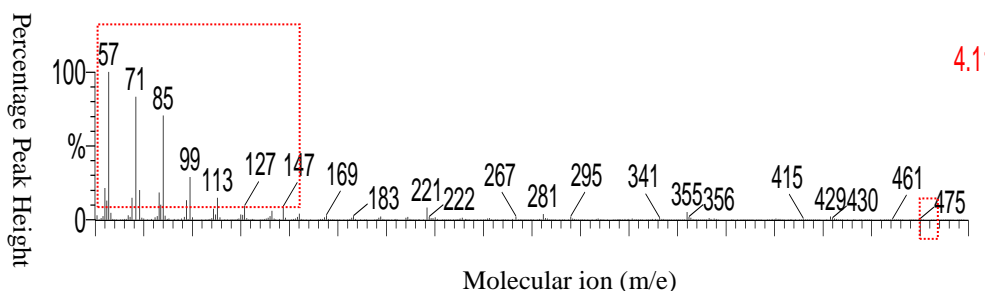
**Figure 5.6** Gas Chromatogram derived from a Fisons 4,000 Gas Chromatography Mass Spectrometer. Internal standard: hexacosane, (C<sub>26</sub>H<sub>54</sub>) retention time 6.46 minutes, showing 100% saturation across the Y axis on the chromatogram.





**Figure 5.7** A Gas Chromatogram was derived from a Fisons 4,000 Gas Chromatography Mass Spectrometer. Internal standard octacosane ( $C_{28}H_{58}$ ), produced 100% saturation at retention time 7.59 minutes. triacontane ( $C_{30}H_{62}$ ) produced a 45% saturation at retention time 9.04 minutes.

Retention time values from figures 5.6 and 5.7, were used for the identification of internal standards, hexacosane, triacontane and octacosane used for quantification techniques. Mass spectra (an example in figure 5.8) were taken across the diameter of the chromatogram peak area to identify the molecular ion content.



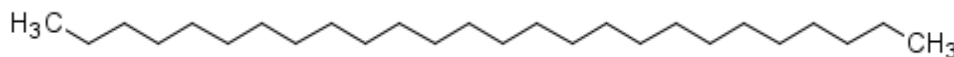
**Figure 5.8** An example of a gas chromatogram electron ionisation mass spectrum. GC-EIMS from within the internal peak area of an internal standard of triacontane, retention time 9.04 minutes (figure 5.7).

Characteristically the mass spectral ions of alkanes typically produce a decreased saturation (figure 5.8), from 57 (100%), 71 (85%), 85 (75%), 99 (40%) and 113 (20%) and 127 (10%). Molecular ion 422 also the molecular weight was also present within figure 5.8, although not visible in this image it was detected using further amplification from within the mass spectra using Mass Lynx software.

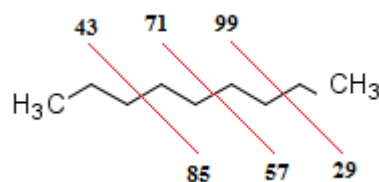
Mass spectral analysis of each gas chromatograph peak was performed by deducting the mass spectra of the background signal from the mass spectra within individual peaks, in order to further identify compounds of interest by mass spectrum signals.

### 5.2.2 Mass spectral fragmentation patterns of alkanes

Fragmentation patterns give a good indication of the molecular structure, a particular molecular ion may represent more than one possible fragmentation event. The whole fragmentation pattern was considered when identifying each substrate. When alkane molecules are fragmented a series of mass spectral ions were detected, known as 'daughter ions'. The 'molecular ion' ( $m/e$  422) usually a prominent large ion, indicated the molecular weight of triacontane, within the Mass Spectrum. The base peak, 100% saturation, is the most stable ion within the mass spectral fingerprint. In nature these stable ions usually occur as; carbonium, allylic, benzylic, aromatic, oxonium and immonium ions. A gap of 14 mass units apart within the triacontane mass spectrum was seen, due to the loss of  $\text{CH}_2$  methyl groups from the carbon chain ( $m/w$  of carbon=12, molecular weight of hydrogen=1).



**Figure 5.9** The figure represents an example of the carbon backbone configuration of alkanes; hexacosane, octacosane and triacontane.



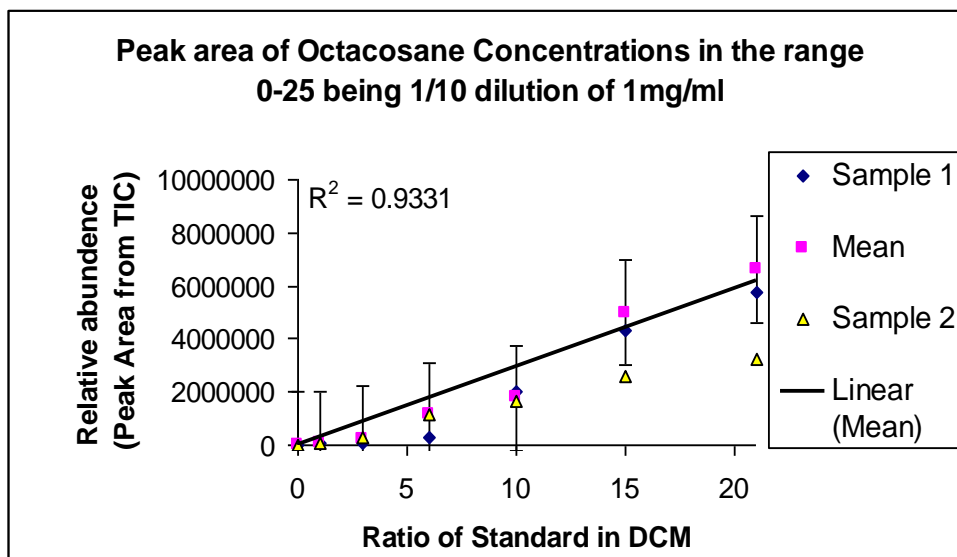
**Figure 5.10** A representative example of fragmentation patterns typically formed from a carbon ( $\text{C}_8$ ) chain.  $\text{CH}_3$  represents a methyl group. Red lines represents the area where cleavage occurs. The upper value indicates mass (m/e) of left hand fragment, whereas the lower value represents the mass (m/e) of the right hand fragment at that point of cleavage.

Fragmentation patterns where the loss of a methyl group is visible, are indicative of straight chain carbon molecules such as 'alkanes'. Signals within the mass spectrum at m/e 69 and m/e 41 are a result of dehydrogenation of m/e 71 and m/e 43 respectively. The greatest signal value within the mass spectra of the alkanes were m/e 57, m/e 71 and m/e 85, which  $[\text{C}_4\text{H}_9]$ ,  $[\text{C}_5\text{H}_{11}]$ , and  $[\text{C}_6\text{H}_{13}]$  respectively.

**Table 5.3** The table lists the name, isotope and atomic weight of small molecules which are constituents of triterpenes. The information was used for derivitisation methods.

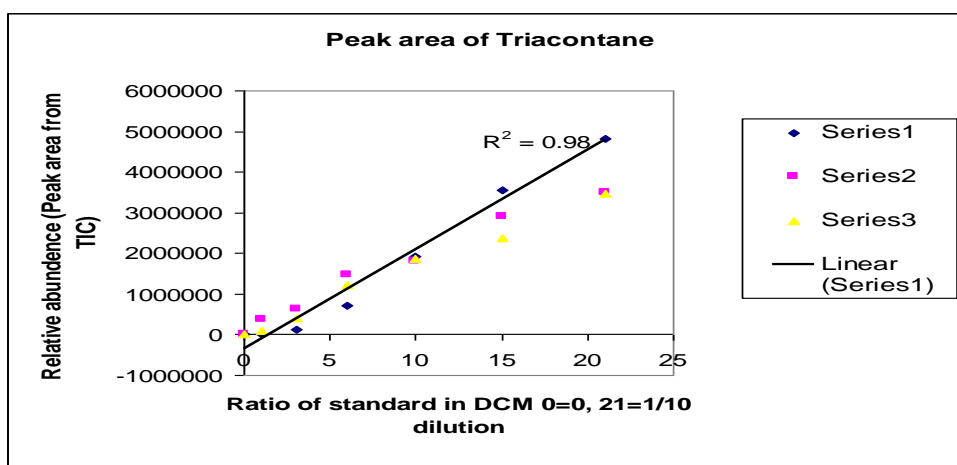
Name	Isotope	Atomic weight
Carbon	C	12
Hydrogen	H	1
Oxygen	O	16
Methyl group	$\text{CH}_3$	15
Water	$\text{H}_2\text{O}$	18
Hydroxyl	OH	17

A calibration curve plotting the peak area ratio against concentration of analyte was developed for use in quantifying analytes. A selected ion remittance chromatogram m/e 85, produced clean detectable peaks.



**Figure 5.11** The total peak area of octacosane isolated by gas chromatography, was measured using Mass lynx software for calibration used for quantification. 1 represented a 1/10 dilution of 1 mg/ml of triacontane.

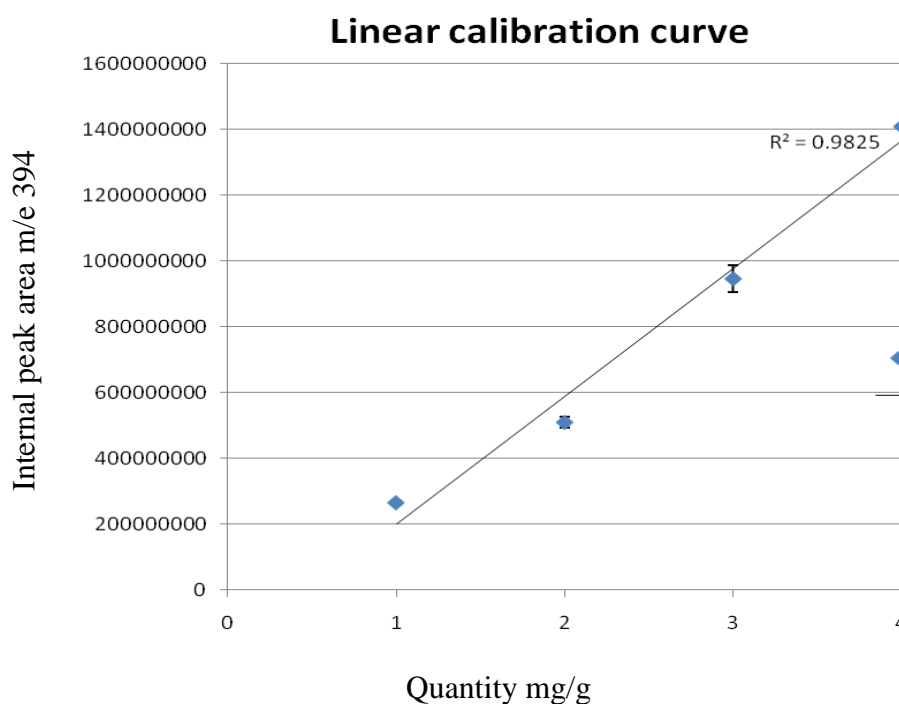
The linear calibration curves of internal standards allowed the single point external standard quantification method to be used. A response factor; the ratio of the signal produced by the gas chromatogram, to the amount of analyte present, was used to quantify metabolites.



**Figure 5.12** The Total peak area values calculated from measuring the internal peak height of the most abundant m/e on the gas chromatogram, were used for calibration of known alkane standards. A dilution ratio of octacosane to dichloromethane was used to produce a linear calibration curve. Series 1, 2 and 3 were repeat calibration measurements.

The formula outlined below was used to calculate the response factor:

$$\begin{aligned}\text{Response Factor} &= \frac{\text{Peak Area}}{\text{Sample quantity}} \\ \text{Quantity of analyte} &= \frac{\text{Peak area}}{\text{Response factor}}\end{aligned}$$



**Figure 5.13** A calibration curve showing the peak area originating from TIC m/e 394 against concentration of octacosane, used for calculating concentration of triterpenes.

The response factor enabled a quantitative value to be determined as a result of peak area values. All peak areas were normalised according to the peak area of the internal standards. The response factor of 392353 was utilised in calculating the concentration of metabolites in mg/g.

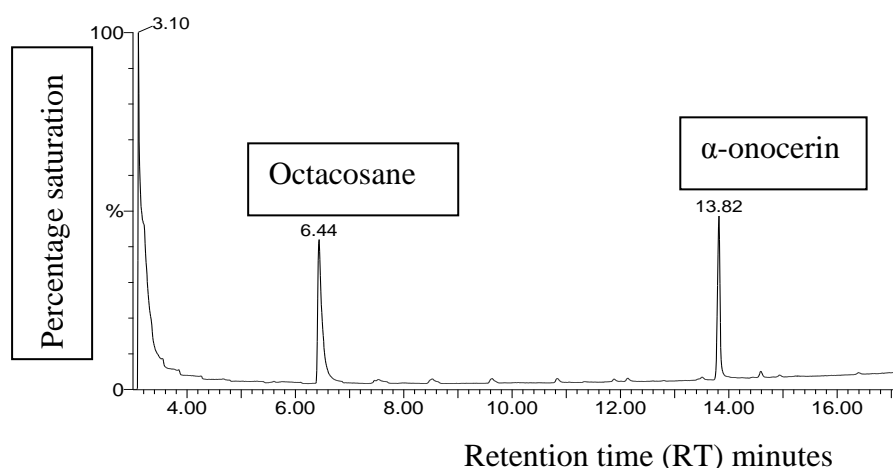
### 5.2.3 Percentage recovery experiment from NSL fractions

A spiking experiment was carried out to determine the percentage recovery of internal standards used for the quantification (triacontane) and external standard ( $\alpha$ -onocerin). The peak area for the largest molecular ion, was measured in triplicate from sample suspended in dichloromethane (1 mg/ml). Three biological triplicates from extracts from *G. glabra*, *A. thaliana*, and *O. pusilla* were spiked with 1 mg/ml of each standard suspended in dichloromethane to test the recovery rate of the extraction method.

**Table 5.4** The table shows the percentage recovery of octacosane and  $\alpha$ -onocerin, at a concentration of 1 mg/ml suspended in dichloromethane. Percentage recovery from dichloromethane and plant extract was measured.

Compound	Species	Tissue	% Recovery from DCM	% Recovery from Plant extract
Octacosane	<i>G. Glabra</i>	Leaf	96	94
	<i>A. thaliana</i>	Leaf	98	96
	<i>O. Pusilla</i>	Leaf	96	96
Octacosane	<i>G. Glabra</i>	Root	92	96
	<i>A. thaliana</i>	Root	97	97
	<i>O. Pusilla</i>	Root	96	96
Onocerin	<i>G. Glabra</i>	Leaf	98	97
	<i>A. thaliana</i>	Leaf	94	93
	<i>O. Pusilla</i>	Leaf	95	95
Onocerin	<i>G. Glabra</i>	Root	96	96
	<i>A. thaliana</i>	Root	96	96
	<i>O. Pusilla</i>	Root	99	95

Octacosane and  $\alpha$ -onocerin standards were suspended in a dichloromethane solution and in whole leaf and root plant extracts from *A. thaliana*, *G. glabra* and *O. pusilla*. The spiked samples were injected into a Fisons gas chromatography mass spectrometer. The peak area of the gas chromatogram from the spiked samples was measured, to assess the recovery rate from non-saponifiable lipid extracts. Percentage recovery rates were summarised in the table 5.4.



**Figure 5.14** A representative example of a Gas Chromatogram showing C<sub>28</sub> Alkane (RT 6.44 minutes) and α-onocerin (RT 13.82 minutes), from an ethanolic/potassium hydroxide extract of *O. pusilla* root tissue. The gas chromatogram shows the percentage saturation and the retention time the triterpenes are separated.

#### 5.2.4 Optimisation and extraction efficiencies from plant extracts: temperature / time experiment

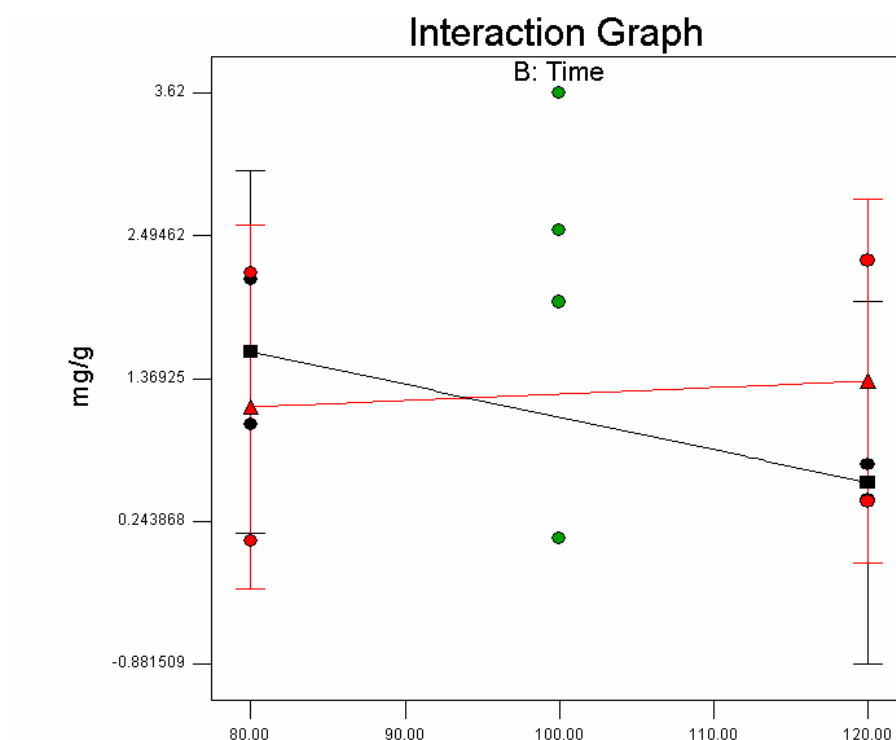
An optimisation experiment to measure the affect of the extraction time and temperature of the incubation period during the non-saponifiable lipid extraction, conducted. The differences between temperature (A) and time (B), parameter values were considered as not significant. The optimum time and temperature condition were selected for future extraction experiments. The resultant protocol utilised a two hour incubation time close to the boiling point of the extraction solvent 80°C (Ethanol 78.4 °C, H<sub>2</sub>O 100°C).

**Table 5.5** Peak areas of C<sub>28</sub> internal standard and  $\alpha$ -onocerin from gas chromatograms produced as a result of the time/temperature experiment were calculated by Mass Lynx 3.2 software. Outlined in the table are the raw data, temperature (°C), time in hours, mean peak area of C<sub>28</sub>, mean peak area of  $\alpha$ -onocerin and the calculated weight of  $\alpha$ -onocerin (mg/g), dry weight.

Temperature (°C)	B:time (hours)	Mean Peak area of C28	Mean Peak area $\alpha$ -onocerin	Mean weight (mg/g)
60	3	173099228	158902	0.09
80	2	20468452	403853	1.97
100	1	14103504	96941	0.69
100	3	15357008	60933	0.40
100	3	15902322	365300	2.30
80	2	20276132	515419	2.54
60	1	24411276	246717	1.01
60	3	26281482	578949	2.20
60	2	21140864	764980	3.62
80	2	156754576	179493	0.11
60	1	24711992	532252	2.15
100	1	21281962	88292	0.41

The data in table 5.5 was used to generate the time/temperature interaction graph using the Design Expert software program. The interaction graph produced non-significant values, as a result, time and temperature values could be interpreted as having limited effect on the extraction efficiency.

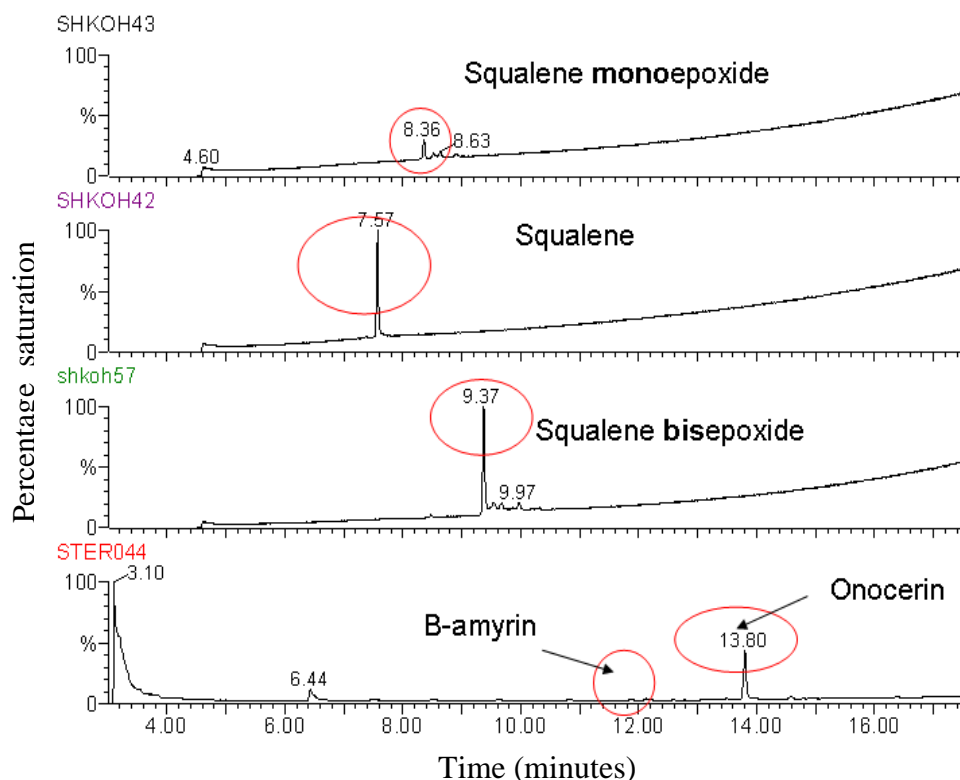




**Figure 5.15** Design Expert interaction graph showing no significant effect of time (minutes) or temperature ( $^{\circ}\text{C}$ ) parameters on the extraction efficiency. From the peak area within the gas chromatogram the total amount  $\alpha$ -onocerin (mg/g fresh weight) was calculated.

### 5.2.5 Mass spectral characteristics and identification of triterpenes


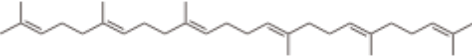
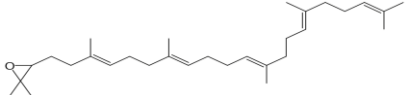
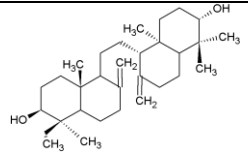
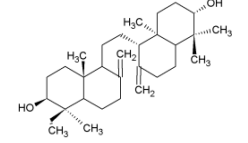
Triterpenes and precursors were identified using; Gas Chromatogram Retention Times (RT), Gas Chromatography Electron Ionisation Mass Spectrometry (GC-EIMS) and Total Ion Chromatograms (TIC). The data generated by these methods were compared with mass spectral data from authentic standards and mass spectral data within the literature.



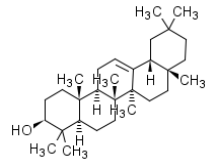
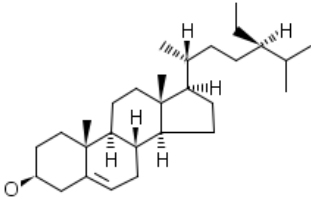
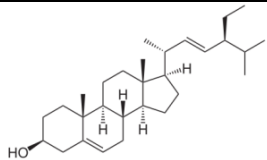
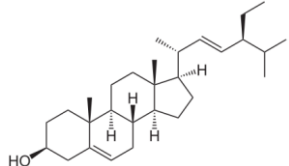
**Figure 5.16** Gas Chromatography Electron Ionisation Mass Spectrometry (GC-EIMS), Total Ion Chromatogram (TIC) showing the retention time of external standards. Squalene (RT 7.57 minutes), squalene monoepoxide (RT 8.36 minutes), squalene bis-epoxide (RT 9.37 minutes),  $\beta$ -amyrin (RT 12.1 minutes) and  $\alpha$ -onocerin (RT 13.80 minutes) suspended in dichloromethane.

The data outlined in figure 5.16 showed that  $\alpha$ -onocerin and precursors, squalene, squalene epoxide (2,3-oxidosqualene) and squalene bisepoxide (2,3,22,23-dioxidosqualene) were able to be separated using Gas chromatography methods developed and outlined in the methods section.  $\beta$ -amyrin was detected as a separate metabolite using Gas Chromatography. The identity of the triterpenes were confirmed using mass spectrum taken from the chromatogram peak areas.

**Table 5.6** The table provides a summary of retention times of triterpene metabolites and standards used for identification; analyte, retention time, molecular ion and molecular formula. A representation of the chemical structure is also shown...

Analyte	Retention time (minutes)	Molecular ion	Molecular formula	Chemical structure
Octacosane	6.46	394	$C_{28}H_{58}$	
Squalene	7.57	410	$C_{30}H_{50}$	
2,3-oxidosqualene	8.36	426	$C_{30}H_{50}O$	
2,3,22,23-dioxidosqualene	9.37	442	$C_{30}H_{50}O_2$	
$\alpha$ -onocerin	13.8	442	$C_{30}H_{50}O_2$	

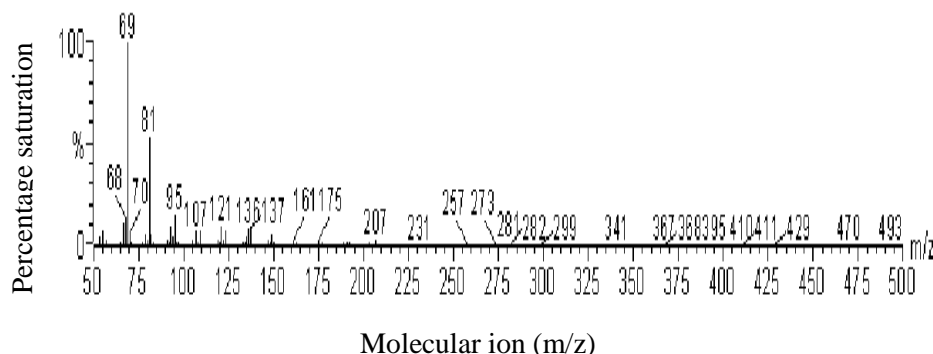
**Table 5.6** continued..... Analytes; octacosane, squalene, 2,3-oxidosqualene, 2,3,22,23-dioxidosqualene,  $\alpha$ -onocerin,  $\beta$ -amyrin, sitosterol, stigmasterol, campesterol.

$\beta$ -amyrin	12.1	426	$C_{30}H_{50}O$	
Sitosterol	10.19	414	$C_{29}H_{50}O$	
Stigmasterol	9.90	412	$C_{29}H_{48}O$	
Campesterol	9.75	400	$C_{28}H_{48}O$	

Mass spectral interpretation tools within Excalibur Qual browser were used for direct comparisons between mass spectra, extracted from total ion chromatograms and selected ion chromatograms. Mass spectra were stacked to allow direct comparisons of individual ions for derivitisation purposes between standards and plant extracts with non-saponifiable lipid (NSL) lipid content.

### 5.2.6 Identification of squalene using GC-EIMS

Squalene was detected in non-saponifiable lipid extracts from plant tissue. A Mass spectrum (figure 5.17), was taken from a peak which occurred at RT 11.79 minutes on a gas chromatogram outlined in figure 5.16. The Mass spectrum was compared directly to a squalene authentic standard (Echelon Biosciences inc., San Jose, USA), and a National Institute of Standards and Technology (NIST) mass spectral library.

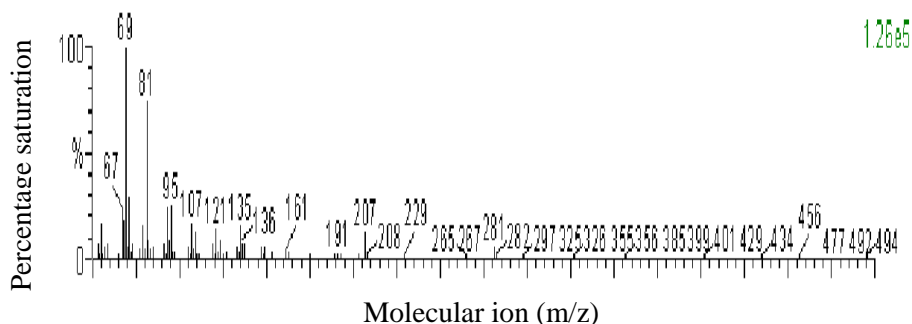


**Figure 5.17** A gas chromatogram-electron ionisation mass spectrum of squalene (RT 11.79 minutes) from a Restharrow NSL plant root extract. The mass spectrum shows the molecular major molecular ions produced by the mass spectrum and the molecular weight (m/w) 410.72.

The structure of squalene in table 5.6 is represented by the fragmentation pattern in figure 5.17. The molecular ions shown in the figure were generated as a result of the fractionation at different points across the compound.

### 5.2.7 Identification of 2,3-oxidosqualene using GC-EIMS

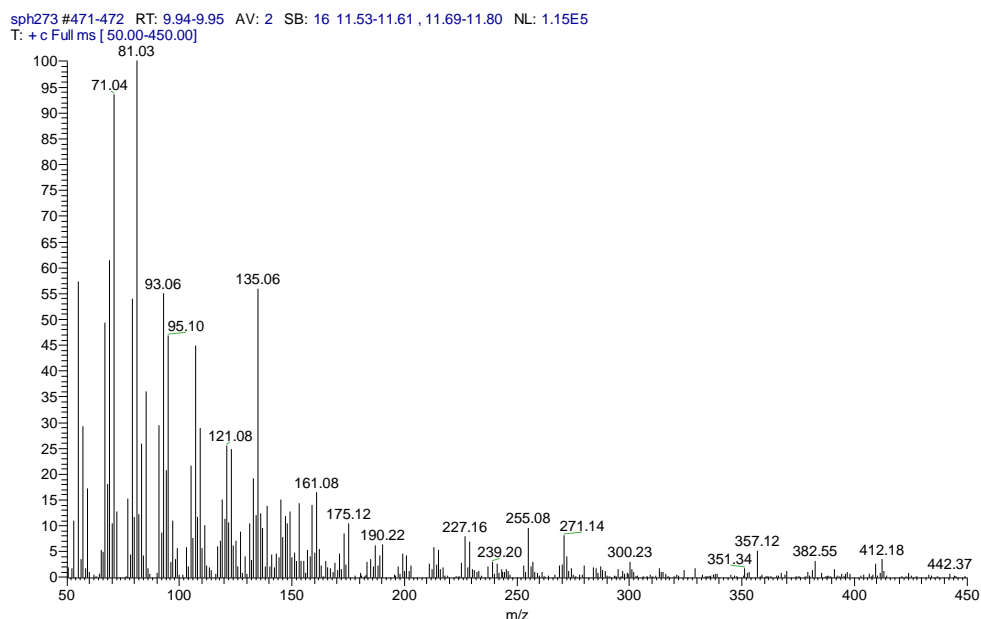
A GC-EI Mass Spectrum of 2,3-oxidosqualene ( $C_{30}H_{50}O$ ) as obtained from an internal standard, suspended in dichloromethane and injected directly into a gas chromatography mass spectrometer. This was directly compared to extracts taken from non-saponifiable lipid plant extracts. 2,3-oxidosqualene had a low molecular weight of 426.717 g/mol. and occurred at retention time of 8.31 minutes (figure 5.16).



**Figure 5.18** A gas chromatogram-electron ionisation mass spectrum of 2,3-oxidosqualene (squalene mono-epoxide), (RT 8.31 minutes), from a Restharrow NSL plant root extract. M/z 69 showing 100% saturation.

### 5.2.8 Identification of 2,3;23,22-dioxidosqualene using GC-EIMS

A GC-EI-mass spectrum of 2,3;22,23-dioxidosqualene ( $C_{30}H_{30}O_2$ ), was compared to analytical grade standards purchased from Echelon Biosciences Inc. The mass spectra from extracted root material matched the internal standard, with a molecular weight of 442. Although 2,3;22,23-dioxidosqualene has an identical molecular weight to  $\alpha$ -onocerin it had a lower retention time than  $\alpha$ -onocerin due to the non-cyclic structure, and allowed separation between the two metabolites by retention time. Mass spectra were taken from each individual sample for confirmation and comparison.

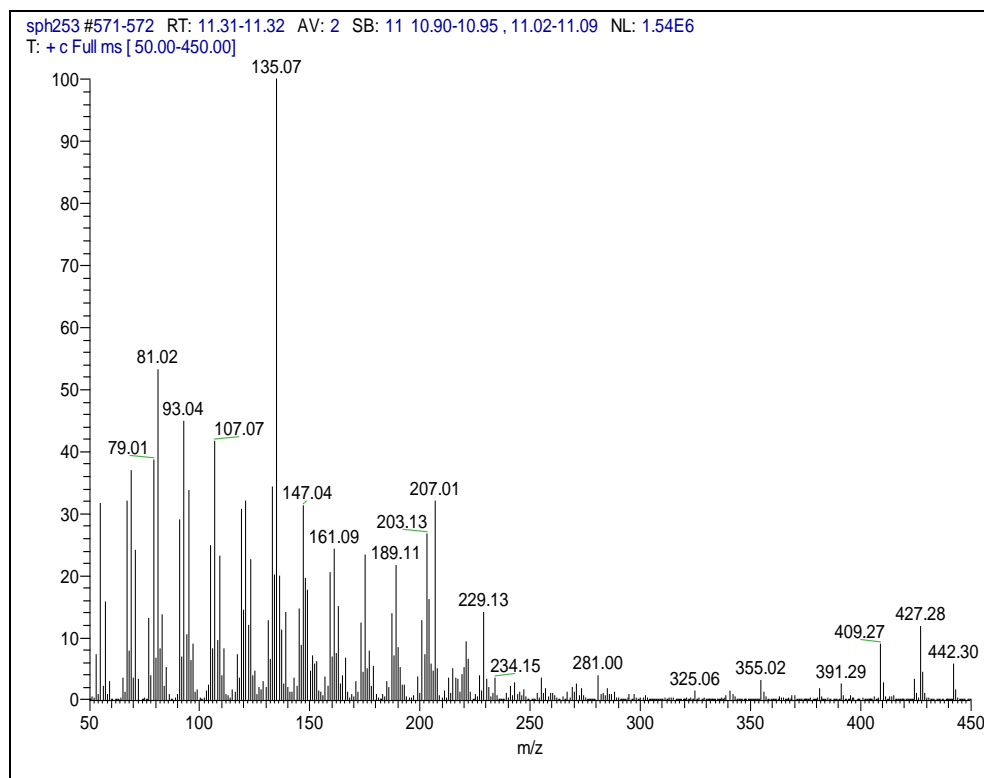


**Figure 5.19** A gas chromatogram-electron ionisation mass spectrum (RT 9.37 minutes) of 2,3;22,23-dioxidosqualene (squalene bis-epoxide), from a Restharrow NSL plant root extract.

A GC-EIMS of 2,3;22,23-dioxidosqualene, was compared to analytical grade standards purchased from Echlon Biosciences Inc. The mass spectra from extracted root material closely matched the internal standard, having a molecular weight of 442.

### 5.2.9 Identification of $\alpha$ -onocerin ( $C_{30}H_{50}O_2$ ) using GC-EIMS

A GC-EIMS spectrum was identified as  $\alpha$ -onocerin. The presence of the molecular ion  $m/e$  442 at less than 10% saturation indicated the compound, had an identical molecular weight to  $\alpha$ -onocerin. The peak at 427 represented a loss of a methyl group ( $CH_3$ ), the peak at 424 represented a loss of water ( $H_2O$ ) and the peak at 409 represented a loss of a water and methyl group. Molecular ion 391 represented the loss of two water molecules and a methyl group, which would have a molecular weight of 51.



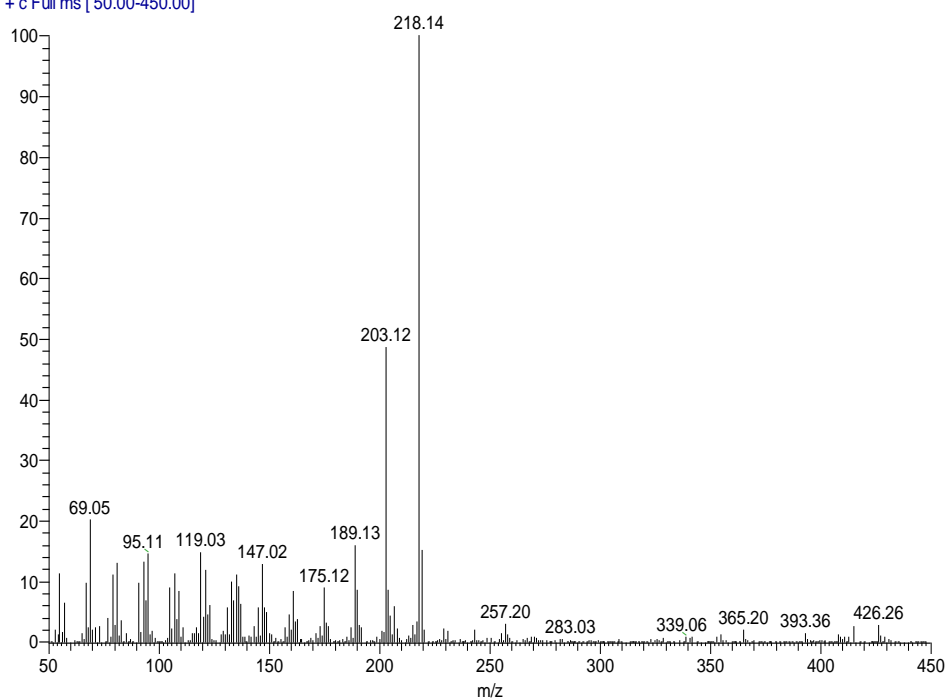
**Figure 5.20** A mass spectrum, of an analyte taken from a gas chromatogram retention time 11.20 minutes. The structure was identified as  $\alpha$ -onocerin from an authentic standard (Peter D.G. Dean, personal communication, 2006).

#### 5.2.10 Identification of $\beta$ -amyrin ( $C_{30}H_{50}O$ ) using GC-EIMS

An authentic standard of  $\beta$ -amyrin was supplied as a gift by Peter, D.G. Dean (2006). The standard was suspended in dichloromethane 10 mg/ml, (w/v) and injected into a gas chromatography electron impact mass spectrometry apparatus using the experimental conditions as outlined in the methods section. The sample was identified as  $\beta$ -amyrin.  $\beta$ -amyrin was subsequently extracted and isolated from plant extracts in both leaf and root material from Restharrow. Mass spectral data taken from total ion chromatograms of the plant extracts, were compared with the authentic standards, producing a molecular ion of 426, typically present in  $\beta$ -amyrin mass spectra. The prominent peak at 218 (100% saturation) and the retention time of 10.42 minutes created a reference point aiding the identification of  $\beta$ -amyrin.



sph254 #505-506 RT: 10.41-10.42 AV: 2 SB: 7 10.47-10.53 , 10.53-10.54 NL: 2.82E5  
T: + c Full ms [ 50.00-450.00]

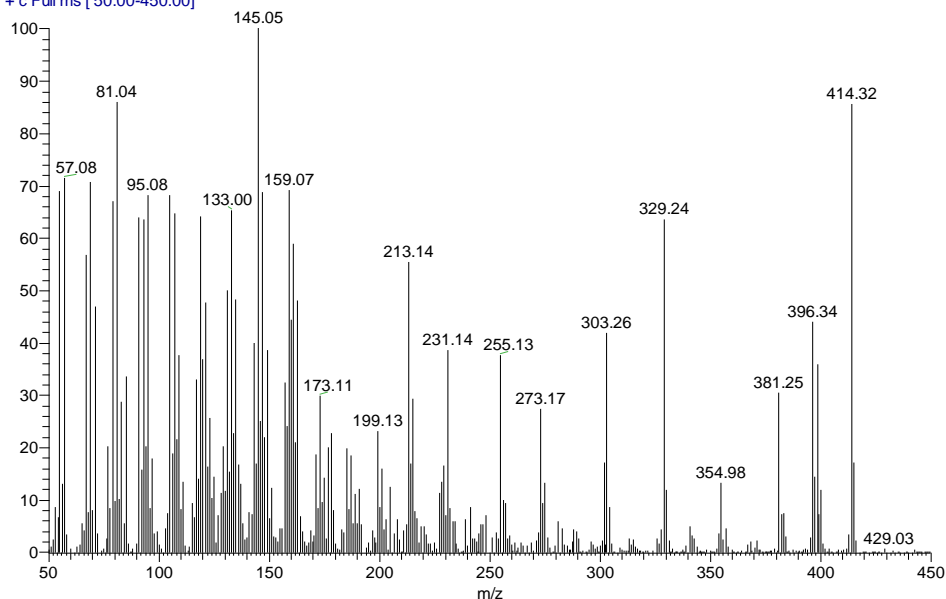


**Figure 5.21** Mass Spectrum of  $\beta$ -amyrin (RT 10.41 minutes), from chromatogram figure 5.16. 100% saturation for molecular ion 218.

### 5.2.11 Identification of sitosterol ( $C_{30}H_{50}O$ ) using GC-EIMS

Sitosterol had a characteristic mass spectrum, which allowed characterisation by comparing with mass spectral data as listed in the literature. Sitosterol had the molecular ion of 414 g/mol, which had a 87% saturation of the mass spectrum. A peak at 145 showed 100% saturation. Sitosterol was extracted from non-saponifiable lipid content of both leaf and root extracts.

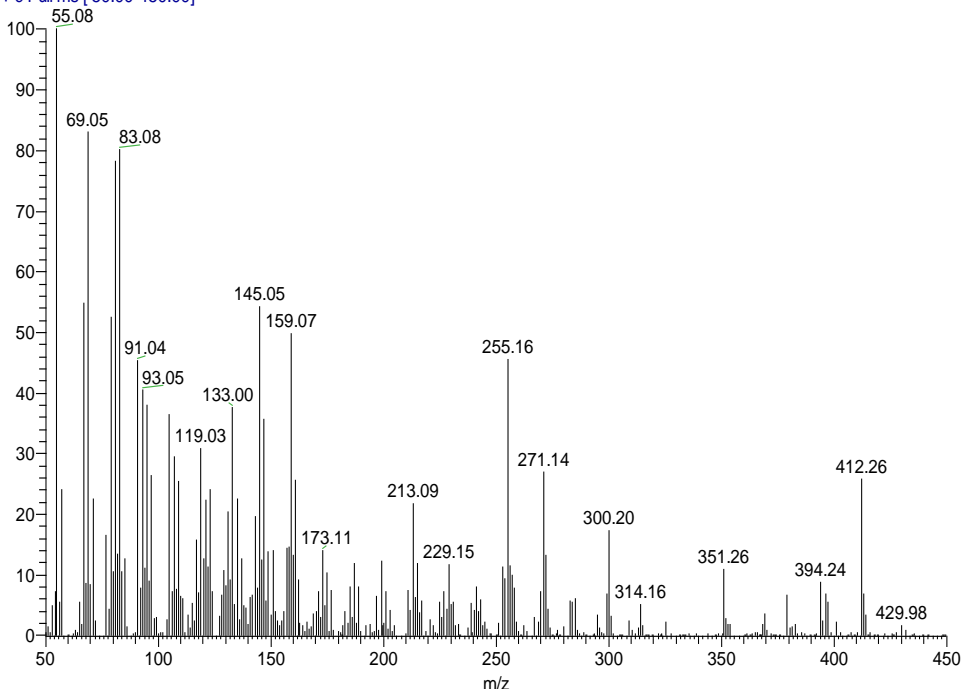
sph255 #489-490 RT: 10.19-10.20 AV: 2 SB: 14 10.00-10.09 , 10.49-10.56 NL: 9.88E4  
T: + c Full ms [ 50.00-450.00]



**Figure 5.22** Gas Chromatography Electron Ionisation Mass spectrum of sitosterol showing relative abundance of ions and their molecular weight. Retention time at 10.19 minutes from plant root extracts figure 5.25.

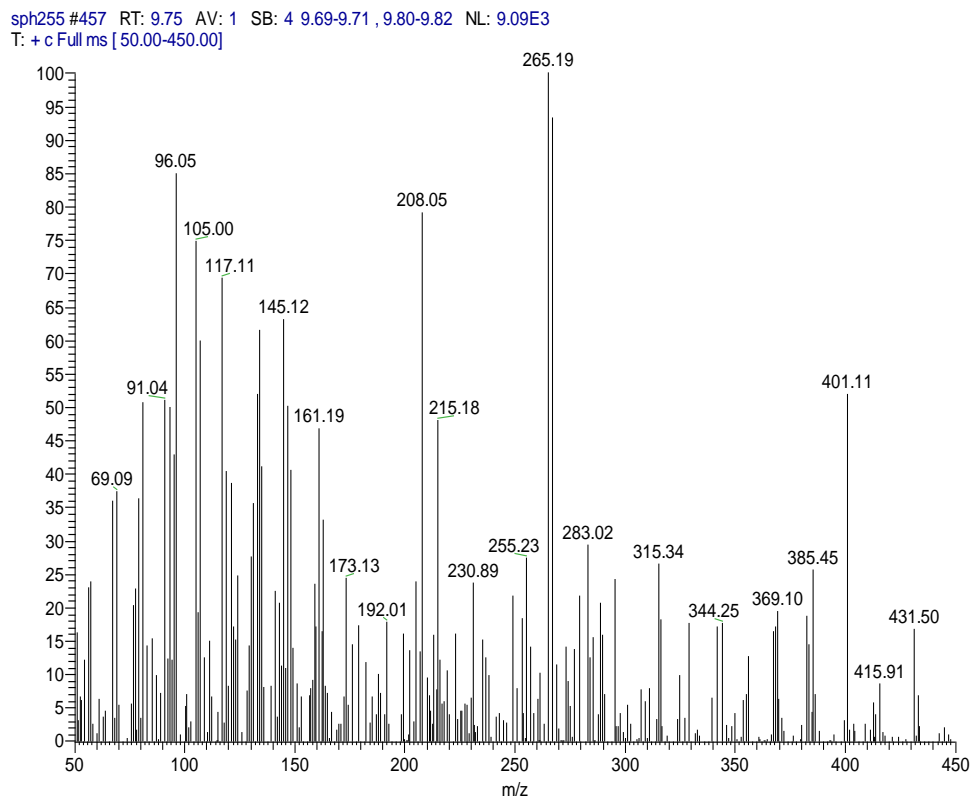
### 5.2.12 Identification of stigmasterol ( $C_{29}H_{48}O$ ), using GC-EIMS

sph255 #468-469 RT: 9.90-9.91 AV: 2 SB: 14 10.00-10.09 , 10.49-10.56 NL: 7.45E4  
T: + c Full ms [ 50.00-450.00]



**Figure 5.23** A GC-EIS mass spectrum from retention time of 9.90 minutes stigmasterol. Molecular ion 412, had a 100% saturation peak at 55, from gas chromatogram 5.25.

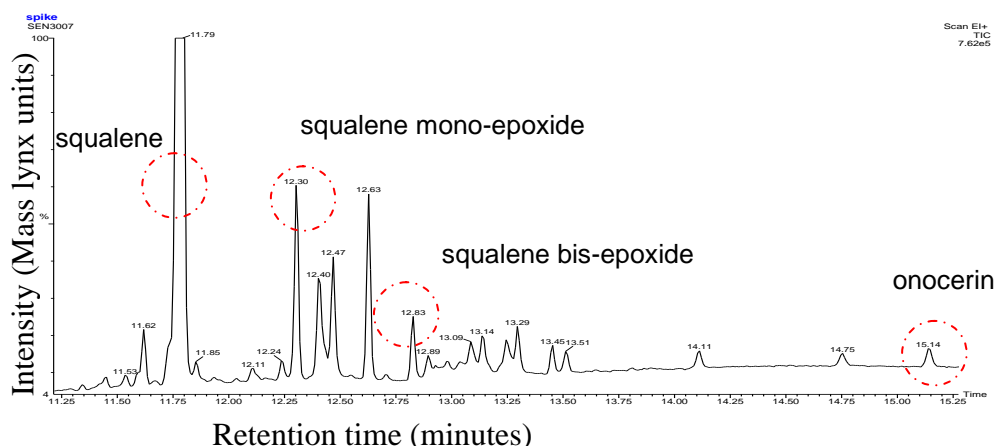
### 5.2.13 Identification of campesterol ( $C_{28}H_{48}O$ ), using GC-EIMS



**Figure 5.24** A gas chromatogram electron ionisation mass spectrum of campesterol as isolated from a NSL extract from Restharrow. Retention time of 9.71 minutes, figure 5.25.

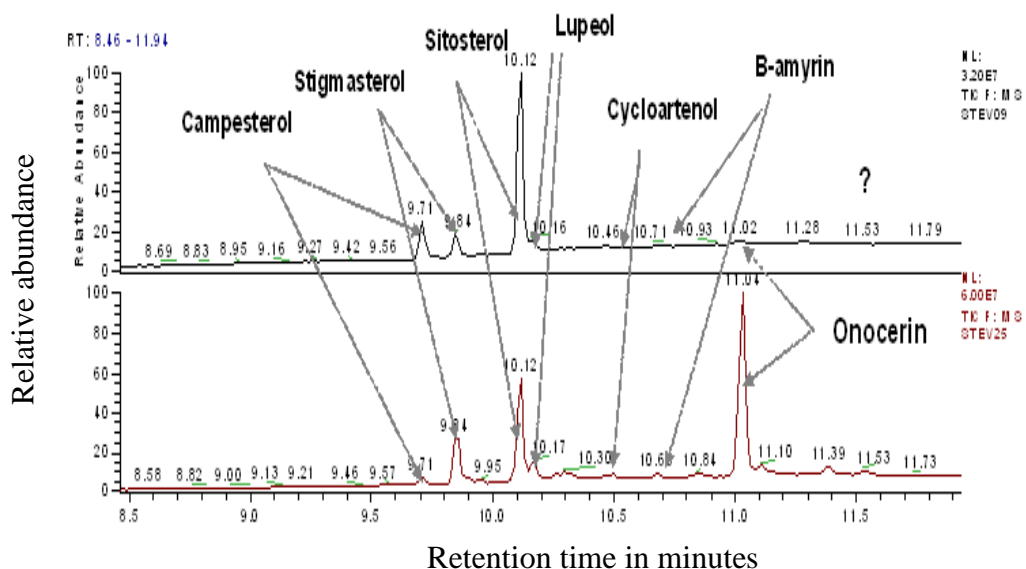
As shown in the GC-EIMS campesterol, an intermediate precursor towards phytosterol biosynthesis was detected in some plant extracts. Campesterol had a molecular ion of 401, detected at retention time of 9.75 minutes from non-saponifiable lipid extracts.

## 5.2.14 Hexane/diethyl ether extraction



**Figure 5.25** A hexane/diethyl ether fraction from *O. spinosa* spiked sample with squalene, squalene mon-epoxide (2,3-oxidosqualene), squalene bis-epoxide (2,3;22;23-dioxidosqualene).

Each compound had a different retention time dependant upon molecular weight. The mass spectra for squalene, squalene mono-epoxide, squalene bis-epoxide and  $\alpha$ -onocerin are all very similar in terms of molecular ions present.

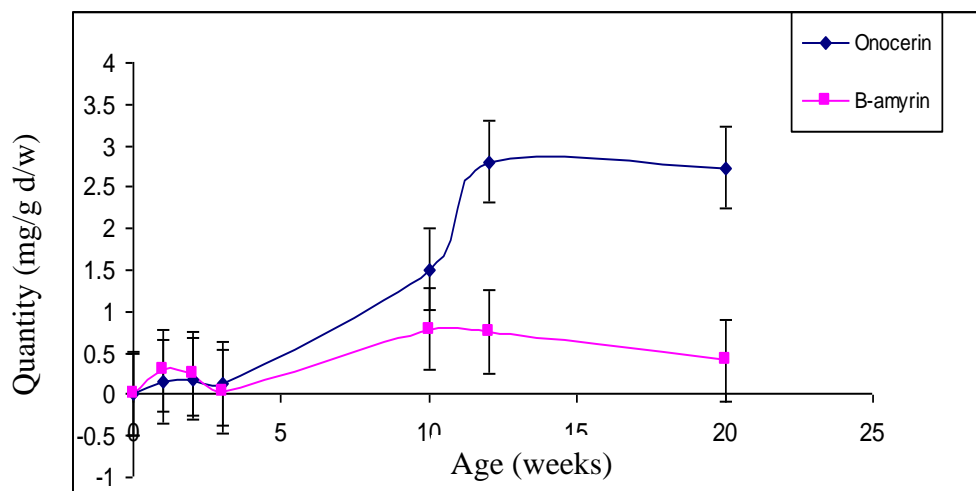


**Figure 5.26** Two gas chromatograms showing NSL hexane/diethyl ether extracts form *A. thaliana* (top) and *O. repens maritima* (bottom).  $\beta$ -amyryn, cycloartenol, lupeol, sitosterol stigmasterol and campesterol are present.

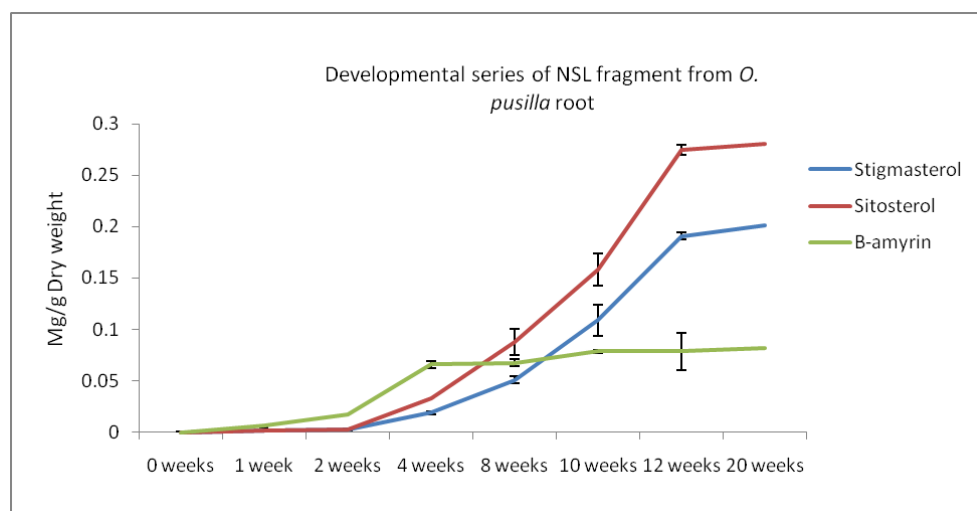
## 5.3 RESULTS: EXPERIMENTAL

### 5.3.1 The occurrence of triterpenoids throughout plant development

An attempt to understand where and when during the plant life cycle  $\alpha$ -onocerin and precursors accumulate was made.



**Figure 5.27** Scatter plot showing metabolite accumulation (mg/g) in root tissue during development in *O. spinosa*.



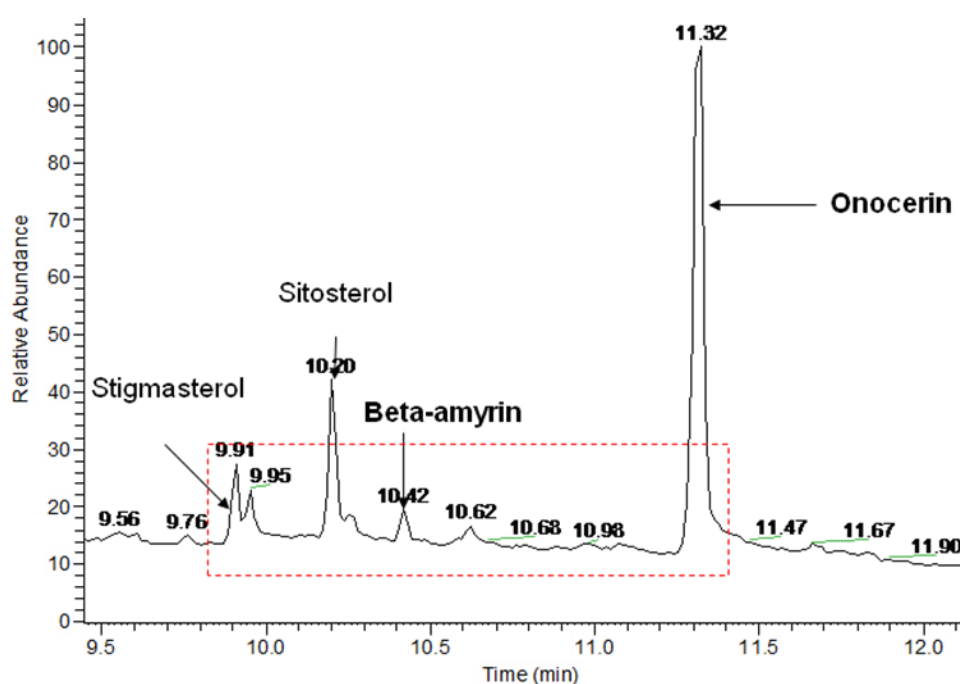
**Figure 5.28** A scatter plot showing metabolite accumulation; stigmasterol, sitosterol and  $\beta$ -amyirin (mg/g dry weight) in *O. pusilla* root throughout development (weeks).

The accumulation of triterpenes was monitored throughout the development of two Restharrow ecotypes; *O. spinosa* and *O. pusilla*. The study focussed on measuring the accumulation of the two secondary metabolites  $\alpha$ -onocerin and  $\beta$ -amyirin shown in figure 5.27. As determined by previous screening and from the literature (Dean and Rowan, 1969), *O. pusilla* has been shown

not to accumulate  $\alpha$ -onocerin at detectable levels. *O. pusilla* was used as a negative control within the experiment, as it was predicted there would be no trade off between the accumulation of either  $\alpha$ -onocerin or  $\beta$ -amyrin. Figure 5.28 showed the accumulation of  $\beta$ -amyrin and phytosterols; sitosterol and stigmasterol throughout development in *O. pusilla* root tissue.

### 5.3.2 Accumulation of triterpenes across Restharrow ecotypes

Non-saponifiable lipid extracts were carried out on six plant species; *O. pusilla*, *A. thaliana*, *G. glabra*, *O. spinosa*, *O. rotundifolia*, *O. repens*, *O. repens subsp. maritima*.



**Figure 5.29** A total ion gas chromatogram from a NSL extract of *Ononis spinosa* root. Phytosterols are shown: stigmasterol (RT-9.91min) and sitosterol (RT-10.20min), along with triterpenes,  $\beta$ -amyrin (RT-10.42min) and  $\alpha$ -onocerin (RT-11.32 min). Retention time in minutes along axis X and relative abundance (peak height) in percentage along axis Y.

Root and leaf tissue was extracted using a diethyl ether/ethanolic extraction method as previously described in Chapter 2 from plants which had reached the 12 week stage. Figure 5.29 shows that stigmasterol (RT-9.91min), sitosterol (RT-10.20min),  $\beta$ -amyrin (RT-10.42min) and  $\alpha$ -onocerin (RT-

11.32 min) were detected in the total ion gas chromatogram from *O. spinosa* root extracts.

### 5.3.3 Multi-species comparison of the triterpenoid metabolome

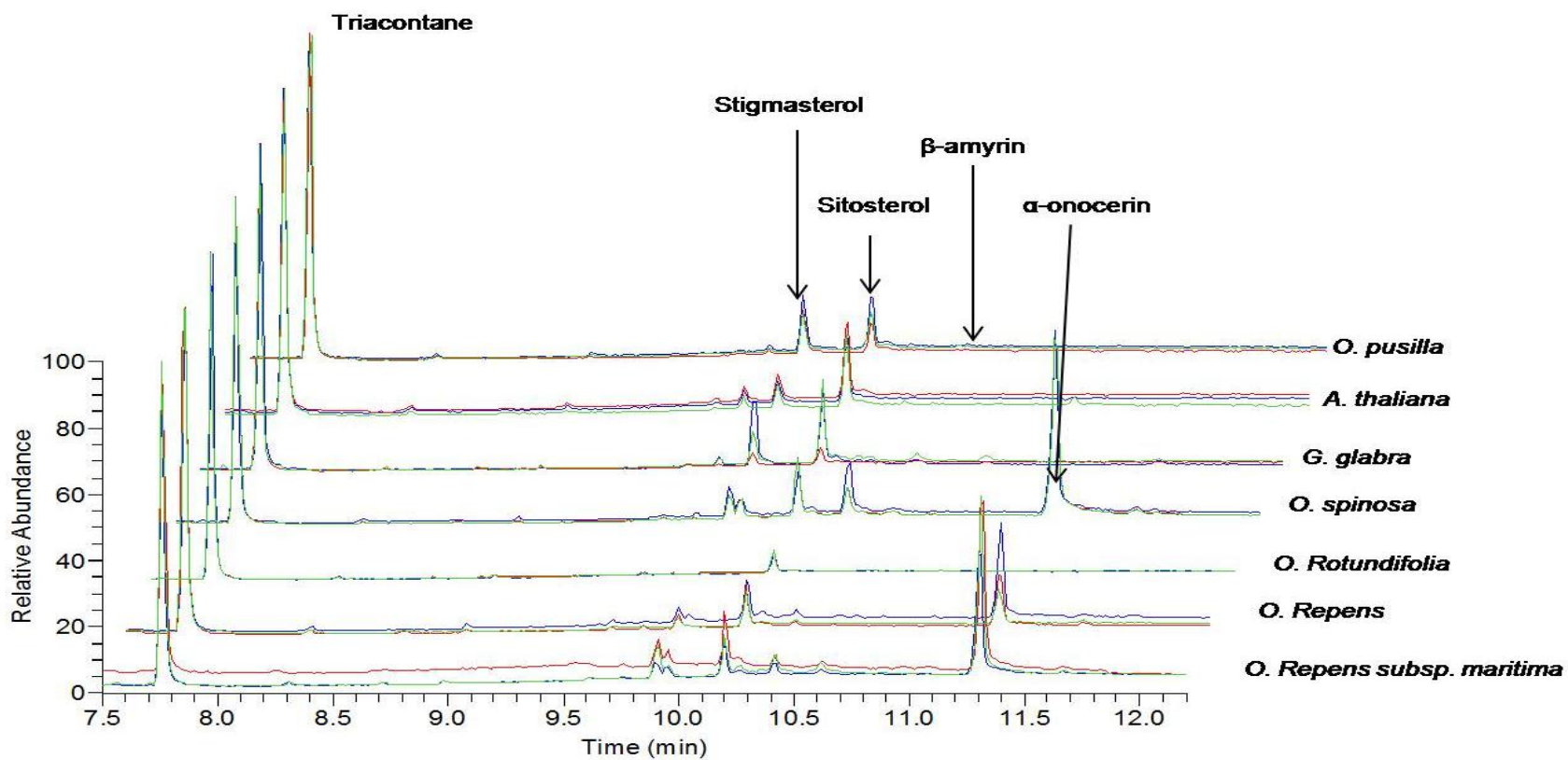
Within the root extracts of all seven species a wide variation of the presence of triterpenes occurred. As expected the amount of  $\alpha$ -onocerin which accumulated showed the largest peak at retention time of 11.32 minutes. As shown in figure 5.30,  $\alpha$ -Onocerin was present in three species of Restharrow: *O. spinosa*, *O. repens* and *O. repens subsp. maritima*. In species *O. pusilla* and *O. rotundifolia* used as within genus controls or *A. thaliana* and *G. glabra*, no  $\alpha$ -onocerin accumulation was detected within the plant root. Other triterpenes detected in the non-saponifiable lipid extract included sitosterol and stigmasterol in all extracts. A peak at retention time 9.37 minutes was identified as 2,3;22-23-dioxidosqualene in the three species where  $\alpha$ -onocerin was also present namely *O. spinosa*, *O. repens* and *O. repens subsp. maritima*.

As shown in figure 5.31, Stigmasterol, sitosterol and  $\beta$ -amyrin were detected from plant leaf extracts from *O. spinosa*, *O. repens* and *O. repens subsp. maritima*. *O. pusilla* and *O. rotundifolia* used as within genus controls or *A. thaliana* and *G. glabra*.

Figure 5.32 shows a comparative representation of NSL extracts between root and leaf. The diagram clearly shows the abundance of  $\alpha$ -onocerin in root extracts of *O. spinosa*, *O. repens* and *O. repens, subsp. maritima*. The diagram also shows  $\alpha$ -onocerin was undetectable in leaf extracts from those species.

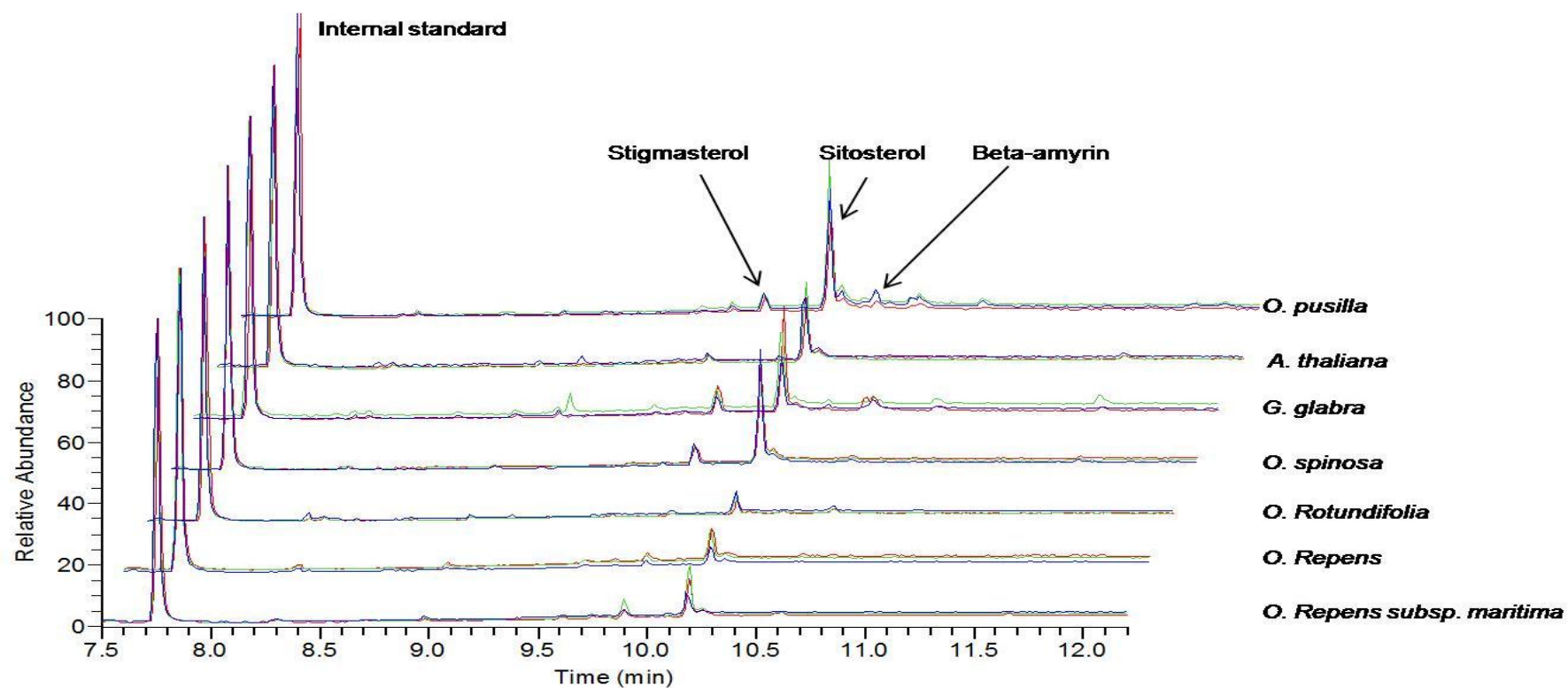
### 5.3.4 Triterpene biochemical profiles within individual species

Species profiles were generated to present a comparative metabolomic content for leaf (figure 5.31) and root tissue for *O. repens subsp. maritima*, *O. repens*, *O. rotundifolia*, *O. spinosa*, *G. glabra*, *A. thaliana* and *O. pusilla*. The concentration of triterpenes in mg/g was compared between species in figures 5.39 (root tissue) and 5.40 (leaf tissue).



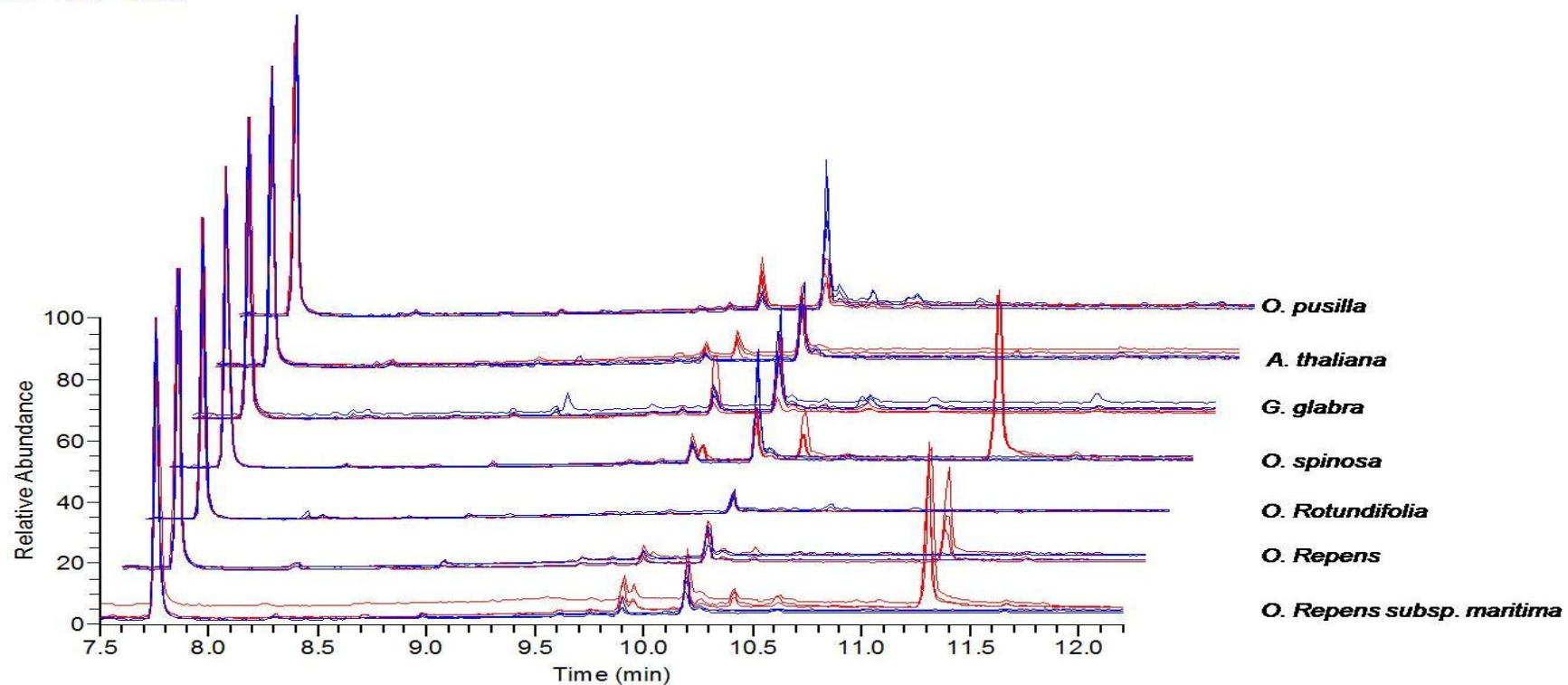
**Figure 5.30** Total ion chromatogram showing relative abundance and retention time (minutes) from non-saponifiable lipid root extracts; biological triplicates from seven different plant species. *O. repens subsp. maritima*, *O. repens*, *O. rotundifolia*, *O. spinosa*, *G. glabra*, *A. thaliana* and *O. pusilla*.



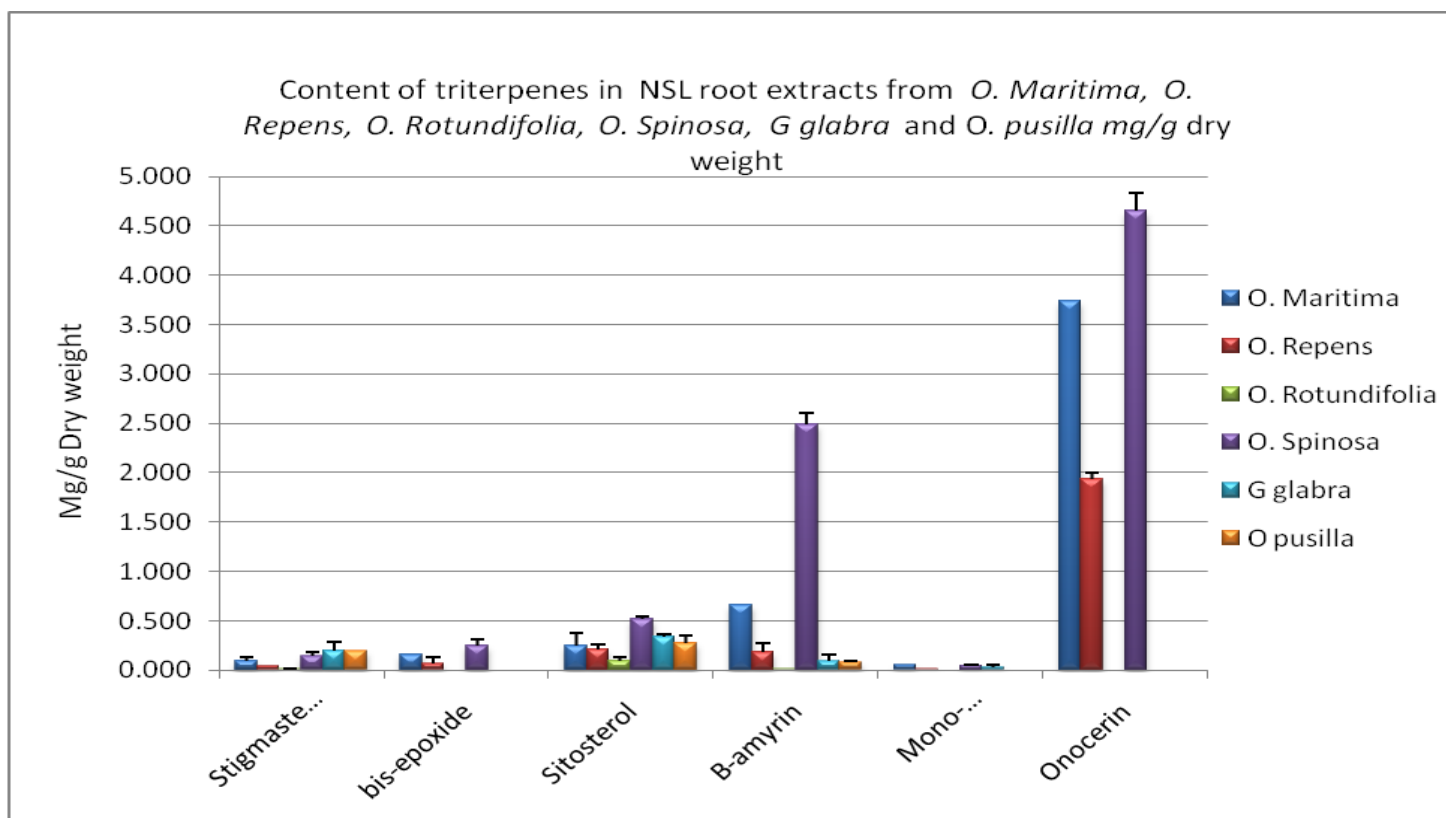


**Figure 5.31** Total ion chromatogram showing relative abundance and retention time (minutes) of NSL leaf extracts of biological triplicates from seven different plant species. *O. repens sub spp. Maritima*, *O. repens*, *O. rotundifolia*, *O. spinosa*, *G. glabra*, *A. thaliana* and *O. pusilla*.

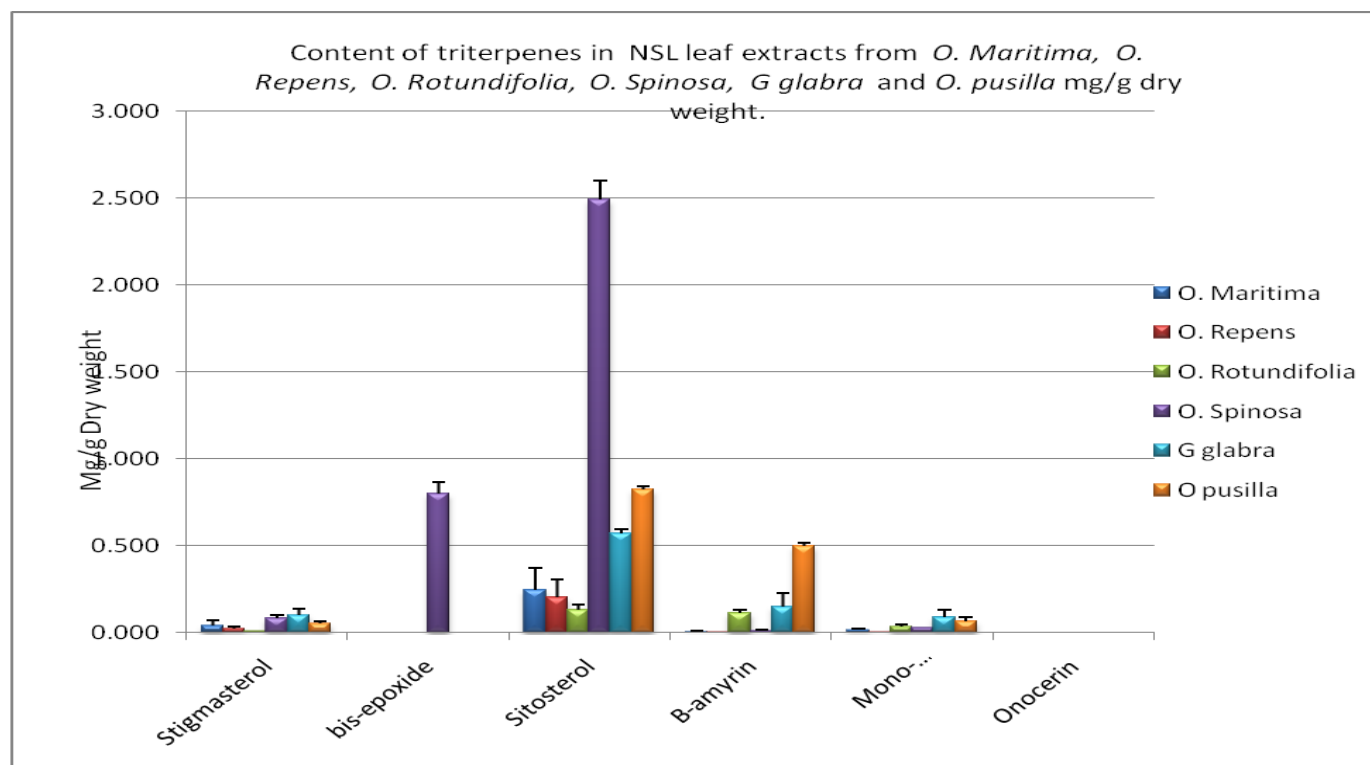
RT: 7.50 - 12.20



**Figure 5.32** Total ion chromatogram showing relative abundance and retention time (minutes) of NSL root and leaf extracts of biological triplicates from seven different plant species. *O. repens sub spp. Maritima*, *O. repens*, *O. rotundifolia*, *O. spinosa*, *G. glabra*, *A. thaliana* and *O. pusilla*. Red = root extract, blue = leaf extract.



**Figure 5.33** Histogram showing, bis-epoxide (2,3;22,23-dioxidosqualene), sitosterol,  $\beta$ -amyryn, mono-epoxide (2,3-oxidosqualene) and  $\alpha$ -onocerin quantities in milligrams per gram dry weight (mg/g d/w). NSL root extracts from *O. repens maritima*, *O. repens*, *O. rotundifolia*, *O. spinosa*, *G. glabra* and *O. pusilla* milligrams per gram dry weight (mg/g d/w).

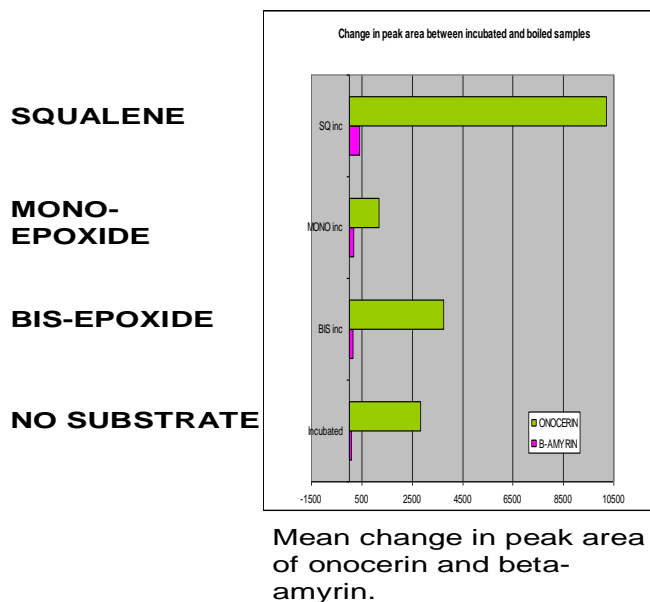


**Figure 5.34** Histogram showing stigmasterol, bis-epoxide (2,3;22,23-dioxidosqualene), sitosterol,  $\beta$ -amyryn, mono-epoxide (2,3-oxidosqualene) and  $\alpha$ -onocerin quantities milligrams per gram dry weight (mg/g d/w). NSL leaf extracts from *O. maritima*, *O. repens*, *O. rotundifolia*, *O. spinosa*, *G glabra* and *O. pusilla* milligrams per gram dry weight (mg/g d/w).

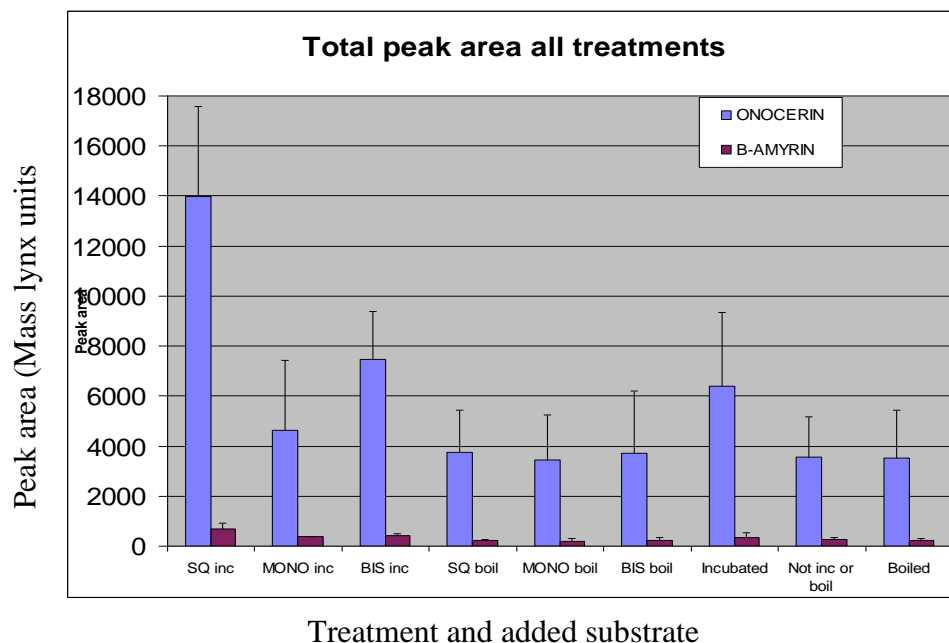
### 5.3.5 In vivo studies of triterpene accumulation in cell preparation of Restharrow: Cell free system For *Ononis spinosa* and cell free system for *O. pusilla*.

A test system was set up to allow comparison of data between non-incubated samples, denatured samples and incubated samples and allow absolute conversion of substrates to be calculated.

An equal volume (200 $\mu$ l) of 10% KOH in 80% ethanol was added to the cell preparation which was vortexed and incubated at 60°C for one hour. After cooling the reaction mixture was extracted three times with 2ml hexane/diethyl ether 10:1. The combined hexane/diethyl layers were washed twice with sterile distilled water. The upper layer of the NSL extract was removed and placed into a new tube. The solvent was evaporated under a stream of N<sub>2</sub> to dryness and re-suspended in 1ml of dichloromethane. Samples were stored at -20°C prior to direct injection. Chromatographic alumina was added to old ether to remove peroxide.



**Figure 5.35** Cell free enzyme incubation experiment. (Left) Expected outcome of substrates being passed through an incubated enzyme substrate complex. (middle) Peak area of products isolated from the enzyme substrate complex. (right) Mean change between  $\alpha$ -onocerin and  $\beta$ -amyrin after incubation.



**Figure 5.36** Cell free enzyme incubation experiment. Total peak area all treatments, boiled, incubated, no-substrate, no incubation, no boil or incubation controls. Substrates added; squalene (SQ), 2,3-oxidosqualene (MONO), 2,3;22,23-dioxidosqualene (BIS).

## 5.4 DISCUSSION

### 5.4.1 Derivatisation of NSL triterpenoids

Methods were developed for the detection of triterpenoid precursors, which may be involved in the  $\alpha$ -onocerin biosynthetic pathway. Squalene (RT 7.57 minutes), 2,3-oxidosqualene (squalene mono-epoxide), (RT 8.36 minutes) and 2,3;22,23-dioxidosqualene (squalene bis-epoxide (RT 9.37 minutes) were all derivitised using Gas Chromatography Electron Ionisation Mass Spectrometry methods. Development of extraction and identification techniques based on comparisons with mass spectral libraries, mass-spectra within the literature, molecular formula calculations (Xcalibur Software), interpretation and inspection of individual ions from within the total ion mass spectrum (TIC) and Selected Ion Mass Spectrums (SIR) provided robust methods to clearly identify secondary metabolites pathway.

Initially, using total ion concentration and select ion transmission scans, the peak area of internal standards were determined. A response factor was calculated from internal standards used in the study. This response factor was utilised in both the normalisation of peak area data and the calculating the final concentrations of metabolites.

A series of analytical chemistry methods were designed and developed for the use in measuring the accumulation of triterpene saponins in both root and leaf tissue within plants. Extraction methods to extract non-saponifiable components were developed to extract components from a multiple stages within triterpene biosynthesis. The successful use of analytical methods offered a reliable and practical approach to measuring phytosterols and triterpene content via the use of gas chromatography mass spectrometry. Details of temperature programs GC settings and mass spectral settings were outlined in the methods section. This information may provide future researchers with a methodology for studies on the ecological function of  $\alpha$ -onocerin in Restharrow.

#### **5.4.2 Developmental series**

The developmental series showed that Restharrow (*O. spinosa*), biosynthesised both  $\alpha$ -onocerin and  $\beta$ -amyirin. There was a trade off between products, initially during early development;  $\beta$ -amyirin was more abundant in *O. spinosa* root. It was interesting that the accumulation of  $\alpha$ -onocerin and  $\beta$ -amyirin was similar at early stages of development, until week 4 and  $\alpha$ -onocerin more prevalent from week 10 onwards. This suggests either a change in transcriptional profile between young and older plants, or change in regulation at the protein level.

#### **5.4.3 Squalene**

Although in the derivitization and method development process squalene was successfully identified and extracted, GC-EIMS failed to detect squalene from plant NSL lipid extracts. The internal standard; octacosane, was applied to root extracts and produced a percentage recovery Table 5.4 from *G. glabra* (96%), *A. thaliana* (98%) and *O. pusilla* (96%). Squalene

may have accumulated in small quantities, beyond the detection limits of the GC-EIMS. Another explanation would suggest that squalene (as the initial precursor towards triterpene and sterol biosynthesis) doesn't actively accumulate in large quantities, being utilised by squalene epoxidases present within the cell. Squalene availability may be a limiting factor in triterpene and phytosterols biosynthesis in plants.

#### **5.4.4 2,3-oxidosqualene**

2,3-oxidosqualene (squalene mono-epoxide) was detected using the molecular ion 426 and was found to be present in all plant species within the arerial fraction; *O. Repens maritima* (0.016 mg/g, d/w), *O. repens* (0.005), *O. rotundifolia* (0.035 mg/g, d/w), *O. spinosa* (0.027 mg/g d/w), *G. glabra* (0.091 mg/g, d/w) and *O. pusilla* (0.067 mg/g d/w). 2,3-oxidosqualene was not present in some of the root extracts, this may be due to the detection limit of CG-EIMS, however accumulation interestingly varies between tissue types. No 2,3-oxidosqualene was detected in *O. pusilla* or *O. rotundifolia*.

#### **5.4.5 2,3;22,23-dioxidosqualene**

2,3;22,23-dioxidosqualene (squalene bis-epoxide) was successfully detected in three Restharrow species. 2,3;22,23-dioxidosqualene accumulated in the roots of *O. maritima* (0.156 mg/g, d/w), *O. repens* (0.063 mg/g, d/w) and *O. spinosa* (0.250 mg/g, d/w). 2,3;22,23-dioxidosqualene was not detected in any of the plant leaf extracts with the exception of in *O. spinosa* (0.800 mg/g, d/w). This is an exceptionally high value, for the presence of 2,3;22,23-dioxidosqualene. There are several explanations which may need to be considered before accepting this value as a true result. The first would be a cross contamination between samples owing to human error. The samples were placed in the GC-EIMS auto-sampler in a random order and standards were run separately, to avoid cross contamination and carryover. The GC-EIMS was run for a period of one hour prior to sampling, to accomplish a column burn eliminating any sample residue from the GC column which may have occurred during previous experiments. The most likely explanation is the automatic peak recognition system accidentally



called a partial peak area of stigmasterol as a 2,3;22,23-dioxidosqualene. By looking at the raw data in figure 5.33 and 5.34 and the original gas chromatogram, it is possible to accept 2,3;22,23-dioxidosqualene is absent from *O. spinosa* leaf extract.

#### 5.4.6 Stigmasterol and sitosterol

Stigmasterol and sitosterol are products of the phytosterol metabolic pathway and are biosynthesised via cycloartenol synthase. Phytosterols are direct competitors with other oxidosqualene cyclase enzymes (OSC) for the 2,3-oxidosqualene substrate. Stigmasterol and sitosterol are subsequently biosynthesised via a number of post cyclic modifications and rearrangements by cytochrome P450 enzymes. Essentially the accumulation of cycloartenol may act as a feed-back mechanism for the production of phytosterols. Stigmasterol and sitosterol were present in all extracts at varying levels.

#### 5.4.7 $\beta$ -amyrin

In root  $\beta$ -amyrin was detected in all six plant species; *O. repens subsp. maritima* (0.652 mg/g, d/w), *O. repens* (0.182 mg/g, d/w), *O. rotundifolia* (0.011 mg/g, d/w), *O. spinosa* (2.494 mg/g, d/w), *G. glabra* (0.091 mg/g, d/w), *O. pusilla* (0.079 mg/g, d/w). The concentration of  $\beta$ -amyrin was highest in root extracts from *O. spinosa*, *O. repens maritima* and *O. repens*. Within leaf the  $\beta$ -amyrin content was greatest in extracts from *O. pusilla* (0.498 mg/g, d/w), *G. glabra* (0.147 mg/g, d/w) and *O. rotundifolia* (0.116 mg/g, d/w). The  $\beta$ -amyrin quantity from within leaf extracts was in *O. spinosa* (0.012 mg/g, d/w), *O. maritima* (0.007 mg/g, d/w) and *O. repens* (0.004 mg/g, d/w).

#### 5.4.8 $\alpha$ -onocerin

$\alpha$ -onocerin is known to occur in the roots of several species of Restharrow as extracted from *O. spinosa* by Hlasiwetz in (1855) and *O. cristata* (5.7 mg/g, d/w), *O. arvensis* (5.39 mg/g, d/w), *O. spinosa* (4.12 mg/g, d/w), *O. repens* (2.28 mg/g, d/w), *O. alopecuroides* (2.79 mg/g, d/w) and *O. mitissima* (1.22 mg/g, d/w) by Rowan *et al.*, (1971). A few other groups

have reported  $\alpha$ -onocerin in Restharrow. Daruhazi *et al.*, (2007), have recently investigated large scale NSL biochemical extraction efficiencies for GC-EIMS detection from *O. spinosa* plants. Other studies surrounding  $\alpha$ -onocerin have involved either chemical characterisation (Barton and Overton, 1955) or synthetic chemical synthesis (Stork *et al.*, 1963; van Tamelen, 1968; Mi *et al.*, 2002; Pauli, 1999; Fröhlich and Pauli, 2000).

In this current study  $\alpha$ -onocerin was reported as being present in three species of Restharrow; *O. maritima*, (3.732 mg/g, d/w), *O. repens* (1.937 mg/g, d/w) and *O. spinosa* (4.651 mg/g (d/w), at a level as described above, similar to data presented by Rowan *et al.* (1971). Data from *O. rotundifolia* root and leaf extracts showed no sign of  $\alpha$ -onocerin. There is no information available in the literature as to whether  $\alpha$ -onocerin accumulates in *O. rotundifolia*.  $\alpha$ -onocerin was not present in NSL extracts from *O. pusilla* which confirmed the results published by Rowan, (1971).

*G. glabra* has been reported in the literature as containing  $\alpha$ -onocerin as a component within the roots (Bisset *et al.*, 2001). Contrary to this  $\alpha$ -onocerin was not detected in NSL fractions taken from *G. glabra* aerial and root tissue.

Rowan and Dean (1972a) determined that  $\alpha$ -onocerin is derived from the triterpenoid pathway and is a product of a squalene enzyme substrate complex.  $\alpha$ -Onocerin is thought to be synthesised by a process infrequently found in nature; cyclisation from both termini of a bis-epoxysqualene (2,3;22,23-dioxidosqualene), (Rowan *et al.*, 1971). The results were consistent with those in the literature (Rowan *et al.*, 1971), from screening several closely related species of Restharrow; which are known to produce  $\alpha$ -onocerin; *O. maritima*, *O. repens* and *O. spinosa* and *O. pusilla* does not accumulate  $\alpha$ -onocerin. Interestingly  $\alpha$ -onocerin accumulated in large quantities within the root of plants which where 2,3;22,23-dioxidosqualene was present in the extracts, namely *O. maritima*, *O. repens* and *O. spinosa*. Likewise in plants which did not accumulate 2,3;22,23-dioxidosqualene,

there was no  $\alpha$ -onocerin detected. The implications of these results will be summarised and discussed further in Chapter 7 (Discussion).

#### **5.4.9 Cell free enzyme reaction**

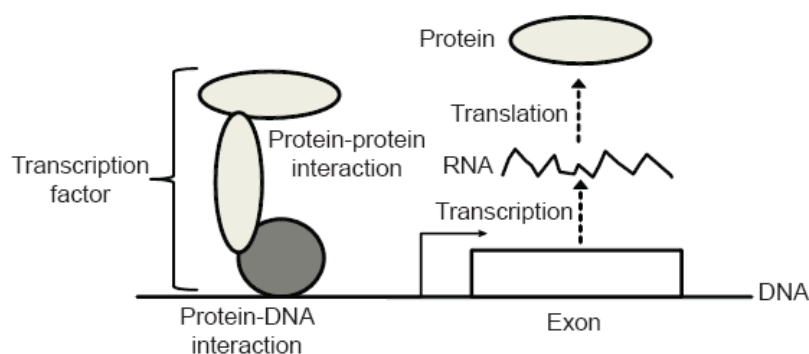
It was noted that from the cell free enzyme incubation studies, the accumulation of  $\alpha$ -onocerin increased when incubated with 2,3;22,23-dioxidosqualene. The main change in  $\alpha$ -onocerin accumulation occurred when the cell free extract was incubated with the addition of squalene. This may indicate that the ability of the plant cell in *O. spinosa* to biosynthesis  $\alpha$ -onocerin is affected at the Squalene epoxidase level rather than at the oxidosqualene cyclase enzyme level. The oxidosqualene cyclase (either  $\beta$ -amyrin synthase or Multi-functional cyclase), may be competing directly for specific substrates. If the primary enzyme involved in  $\alpha$ -onocerin biosynthesis was a squalene bis-epoxidase, it would be expected that the largest change in  $\alpha$ -onocerin accumulation may occur when a cell free enzyme extract was incubated with squalene as a raw substrate. In support of this evidence the accumulation of  $\beta$ -amyrin was seen to accumulate in greater quantities where  $\alpha$ -onocerin was present than in plant which were shown not to accumulate  $\alpha$ -onocerin. This may indicate that a single oxidosqualene cyclase enzyme is active and is substrate dependant.

## CHAPTER 6. X SPECIES TRANSCRIPTOMICS OF *O. SPINOSA* VS *O. PUSILLA*: GENES INVOLVED IN TRITERPENE BIOSYNTHESIS

### 6.1 INTRODUCTION

#### 6.1.1 Transcriptional profiling

The ‘Central Dogma’, while many exceptions exist, forms the basic model for molecular biology. The model outlined in figure 6.1 can be described as, the route from genes at the DNA level, through RNA to protein. For protein coding genes in particular, each gene is transcribed into RNA, the non-coding introns are removed and the mRNA is transported into the cytosol where the mRNA is translated into protein on the ribosomes.



**Figure 6.1** The central dogma of eukaryotic regulatory mechanism (Crick, 1970). The transcription from the DNA exon location, to RNA followed by the translation to protein.

The discovery of transcription factors has been offered as an explanation to describe some discrepancies between mRNA gene expression values and protein levels. There is evidence that micro RNA may be involved in gene regulation, (Anderson and Selthamer, 1997) and that miRNA may control 30% of protein coding genes.

Quantification of large numbers of messenger RNA (mRNA) transcripts using microarray technology can provide a detailed insight into cellular processes involved in the regulation of gene expression. Large-scale

generation of gene regulation information was not possible until recently. Microarray technology was developed and used on a small scale in the mid 1980s (Augenlicht *et al.*, 1987), and did not become available for analysing large transcriptomes until the 1990s (Lockhart *et al.*, 1996). Pioneering work on plant microarrays was presented a decade ago. Microarray experiments are extremely powerful and provide information on a genome-wide scale. However, they are also very complex and can be time-consuming to analyse data compared to other biological techniques.

There are several alternatives to cDNA microarray technology for investigating gene expression. Spotted oligonucleotides allow more gene-specific measurements using a sequence-based platform. Affymetrix has developed a method which allows probes to be synthesised through a photolithographic process *in situ*, allowing a robust platform creating reproducibility between experiments. Quantitative real time reverse transcription PCR (qRT-PCR; Bustin, 2002), is often used for smaller gene number expression comparisons. For larger scale studies, Expressed Sequence Tag (EST) library construction has been widely used for array construction and gene expression profiling (Sterky *et al.*, 2004). The high cost of sequencing has led to other alternatives, such as Serial Analysis of Gene Expression (SAGE) and Massive Parallel Signature Sequencing (MPSS), (Brenner *et al.*, 2000), which both aim to maximise the amount of information extracted per sequence run.

Techniques based on *de novo* sequencing (454 and Alexa™) have allowed increased high-throughput profiling at a feasible cost. The term microarray will refer to spotted cDNA microarrays unless stated otherwise. The generation of short sequences (using Illumina or SOLiD systems) represents a true alternative to microarrays, where there is genome or transcriptome sequence available to assemble short sequence oligonucleotides.

Commercial software for microarray analysis, such as GeneSpring (GX Inc.) and GenePix, have played an important role in providing simple and extensive tools for genetics laboratories. Expression viewer tools are

available as online bioinformatic interfaces: TIGR multi expression viewer provides many clustering and classification algorithms in user-friendly software. The Bioconductor suite mainly written in the statistical language R, offers an extensive collection of methods and algorithms (Gentleman *et al.*, 2004).

### **6.1.2 Cross species transcriptional profiling**

Cross species transcriptional profiling has recently been developed. Examples in animal include; human versus monkey, human versus mouse, within *Xenopus* within *Drosophila* and between rodents. Studies have shown that cross species hybridisation affects neither the reproducibility or the distribution of oligonucleotide probe hybridisation using the Affymetrix, human GeneChip® technology, commercially available (Affymetrix, Santa Clara, USA). Cross species transcriptional profiling has been widely used and extensively validated. Within this technology platform, probe sets consist of 11 or 20 probe pairs, used to quantify the abundance for each transcript. Each probe pair consists of a perfect match (PM) and a mismatch (MM) oligonucleotide. The PM probe consists of a 25 base pair sequence complimentary to the target sequence. The MM probe has a mismatch at the 13<sup>th</sup> base, compared to the PM probe. The hybridisation efficiency of each probe can be derived from the signal differences between the PM and MM oligonucleotides, across the whole probe set (Hammond *et al.*, 2005).

GeneChip® technology has been widely adopted by the plant science research community, with large data sets available from reproducible experiments conducted on the model crop *A. thaliana* (Craigon *et al.*, 2004). At present due to the cost, time and effort involved in producing fully sequenced genomes, only commercially important crops considered as food sources are fully sequenced or are considered model species. As sequencing would be required prior to developing a custom GeneChip® array, there is a gap in the availability of GeneChip® arrays for less extensively studied organisms.

One solution to this problem is to use GeneChip® arrays designed for model organisms to study related species. Whilst this strategy is feasible, studies conducted to date have not accounted for the probability that there will be inefficient hybridisation of transcripts from the target species to GeneChip® probes designed to the other species, due to sequence changes that have occurred since the phylogenetic separation of the two species from a common ancestor. To address the potential problems of using GeneChip® arrays designed for one species to monitor the transcriptome of a closely related species, Hammond *et al.*, (2005) developed a novel technique to improve the sensitivity of high-density oligonucleotide arrays when applied to samples from heterologous species. This technique is based on selecting probe-pairs based on the hybridisation efficiency of the PM oligonucleotide probe with genomic DNA from the target species and assesses the degree of cross-hybridisation to the available GeneChip. Initially the work was carried out using the *A. thaliana* ATH1-121501 GeneChip® array to study the transcriptome of *Brassica oleracea* L. Probe mask files increased sensitivity of the ATH1-121501 GeneChip® array when monitoring the regulation of gene expression in *B. oleracea* under Phosphate stress by up to 13-fold.

Cross species transcriptomics has been utilised a number of times, for identifying the transcript levels in non-model plant species. Hammond *et al.*, (2006) developed a method to compare two transcriptomes from *Thalspi caerulescens*, a hyper accumulator and *Thalspi arvense*, a non-hyperaccumulator. Large scale cross species transcriptomics has been utilised by Broadley *et al.*, (2008), who quantified the expression of 18,494 genes across 14 taxa of Brassicaceae clade. Methods have been developed to test transcriptional changes and cross hybridisation efficiency in distantly related plants. Davey *et al.*, (2009), used the Affymetrix Rice oligonucleotide GeneChip® to profile the response of the banana (*Musa spp.*) leaf transcriptome, when subjected to drought stress.

## 6.2 RESULTS

### 6.2.1 Genomic DNA hybridisation and probe-selection

The GeneChip array *A. thaliana* ATH1-121501 was used as a model to study the transcriptional differences between *O. pusilla* and *O. spinosa*. Polymorphisms between the original *Arabidopsis* oligonucleotide probe sequences, were likely lead to an underestimation of the transcript abundance if all the probes were used. Initially the gDNA from both species was biotin labelled and hybridised to the ATH1 GeneChip array. Probe sets for PM and MM were selected for further analysis above an appropriate threshold. This threshold was calculated using a parser script (NASC Arrays) written in Pearl software programming language (XspeciesVersion1.1; <http://affymetrix.arabidopsis.info/xspecies>).

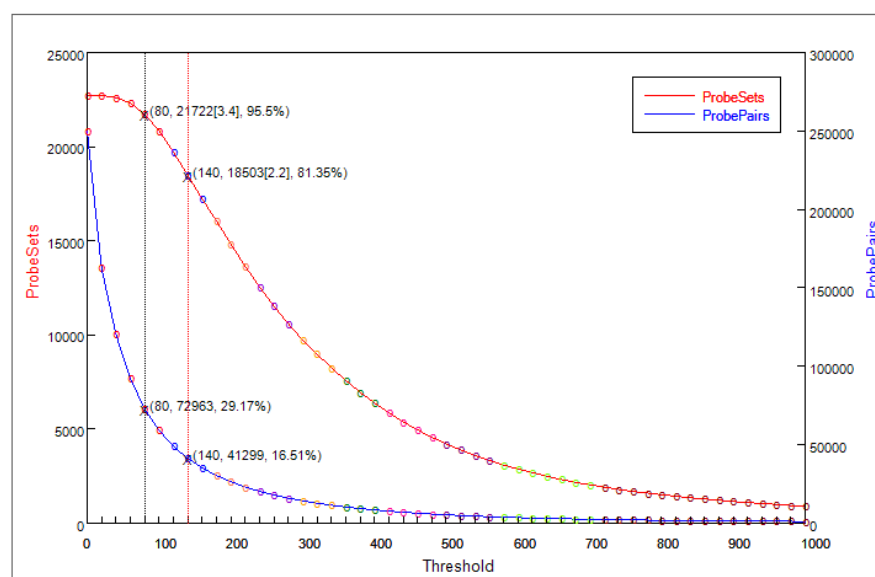
### 6.2.2 gDNA hybridisation intensities

Genomic DNA probe mask files generated from intensities 20-1000 were tested for suitability to analyse the *Ononis* transcriptome on GeneChip ATH1.

Subsequently individual probe sets were selected based upon the hybridisation efficiency of gDNA from *O. pusilla* and *O. spinosa* on ATH1 for expression analysis.

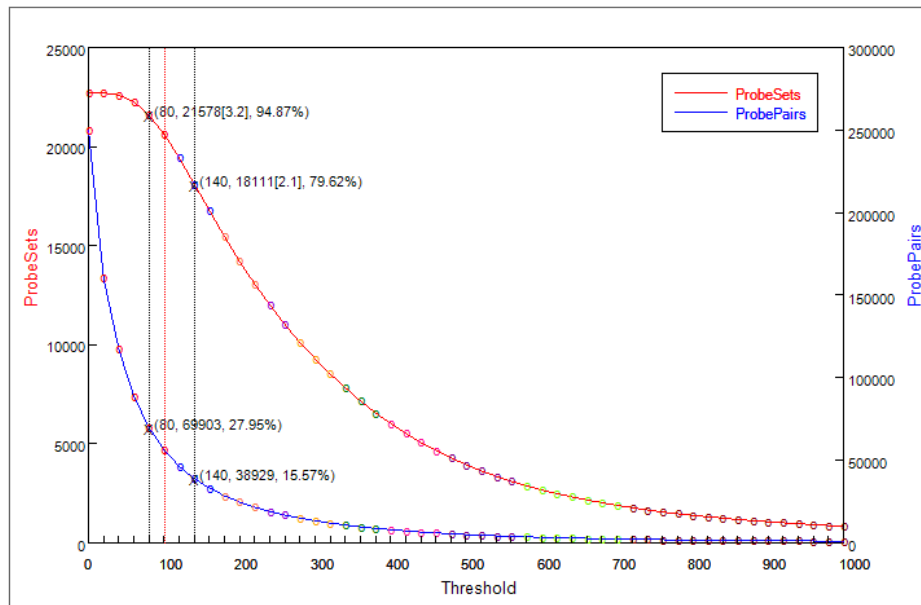
*O. spinosa* gDNA hybridisation signal value was normalised using Robust Means Analysis (RMA) with cross species hybridisation discovery software PIGEON (unpublished, Mayes and Lai (2010)). The threshold of probe pairs against probes sets lost at different thresholds 20-1000 intensity was plotted. For *O. spinosa* gDNA at threshold 100 was used to generate a masking file, also using PIGEON.





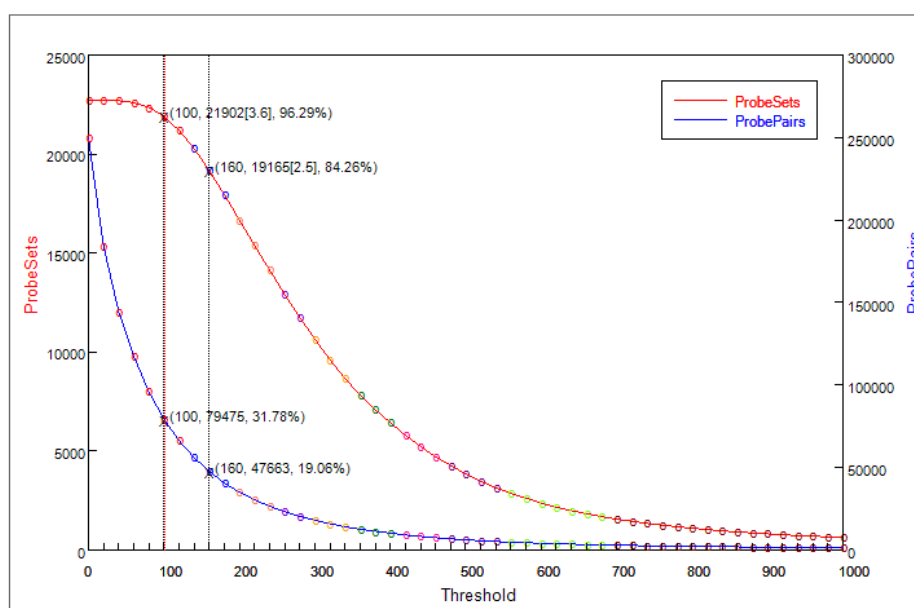
**Figure 6.2** Number of *Arabidopsis thaliana* probe-sets and probe-pairs from the ATH1 – 121501 GeneChip® array retained after gDNA hybridisation using *O. pusilla* at a range of signal cut-off thresholds. (unpublished Mayes and Lai 2010). Selection of an appropriate threshold was used to generate a CDF probe mask file which can be used for transcriptional profiling. The left axis (red trend line) shows the number of probe-sets retained (used in mask files) and the right axis (blue trend line) shows the number of probe-pairs retained (used in mask files) at gDNA hybridisation intensities 20-1000.

At this threshold the number of probe sets retained was 20544 and probe pairs was 55543. Thus 90.75% of probes were retained and 22.61% of probe pairs were retained. This was considered an acceptable level of probe pair loss in comparison to the number of probes retained. For hybridised gDNA, *O. pusilla* retained 20567 (90.41%) probe sets and 56500 (22.59%) probe pairs at a signal threshold of 100.



**Figure 6.3** Number of *Arabidopsis thaliana* probe-sets and probe-pairs from the ATH1 – 121501 GeneChip® array retained after gDNA hybridisation using *O. spinosa* at a range of signal cut-off thresholds. (unpublished Mayes and Lai 2010). Selection of an appropriate threshold was used to generate a CDF probe mask file which can be used for transcriptional profiling. The left axis (red trend line) shows the number of probe-sets retained (used in mask files) and the right axis (blue trend line) shows the number of probe-pairs retained (used in mask files) at gDNA hybridisation intensities 20-1000.

The signal intensity values and the probe sets retained versus the probe pairs removed, was tested using PIGEON. The absolute hybridisation intensities were initially RMA normalised within PIGEON software. The number of probe pairs and sets retained was plotted against threshold values ranging from 0-1000. At a threshold of 100, 22354 (98.28%) probe sets and 96042 (38.4%) probe pairs were retained.



**Figure 6.4** Number of *Arabidopsis thaliana* probe-sets and probe-pairs from the ATH1 – 121501 GeneChip® array retained after gDNA hybridisation using *A. thaliana* at a range of signal cut-off thresholds. (unpublished Mayes and Lai 2010). Selection of an appropriate threshold was used to generate a CDF probe mask file which can be used for transcriptional profiling. The left axis (red trend line) shows the number of probe-sets retained (used in mask files) and the right axis (blue trend line) shows the number of probe-pairs retained (used in mask files) at gDNA hybridisation intensities 20-1000.

### 6.2.3 Cross hybridisation analysis across four plant species

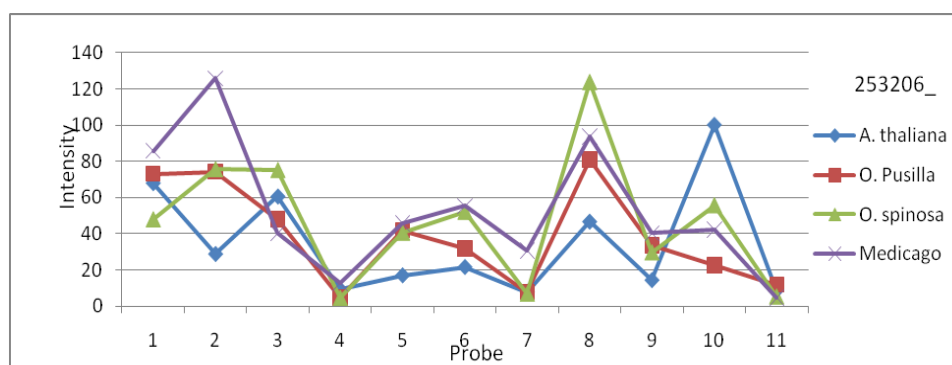
Signal values from hybridised gDNA *O. spinosa* and *O. pusilla* and *Medicago truncatula* on Affymetrix GeneChip *A. thaliana* ATH1 (22K) were normalised using Robust Means Analysis (RMA) PIGEONS software (unpublished, Mayes and Lai (2010)). The signal values for Perfect Match (PM) and MisMatch (MM) were transferred to Excel (.cel) files where the values were compared between species. The hybridisation efficiency was measured and compared primarily between *A. thaliana*, *M. truncatula*, *O. pusilla* and *O. spinosa*. The AGI codes (Arabidopsis Genome Initiative numbers), listed as the Organism (At=*A. thaliana*), the chromosome number

followed by the gene location on the chromosome. The probe IDs (ATH-1 GeneChip® identification numbers) used are listed in table 3.1.

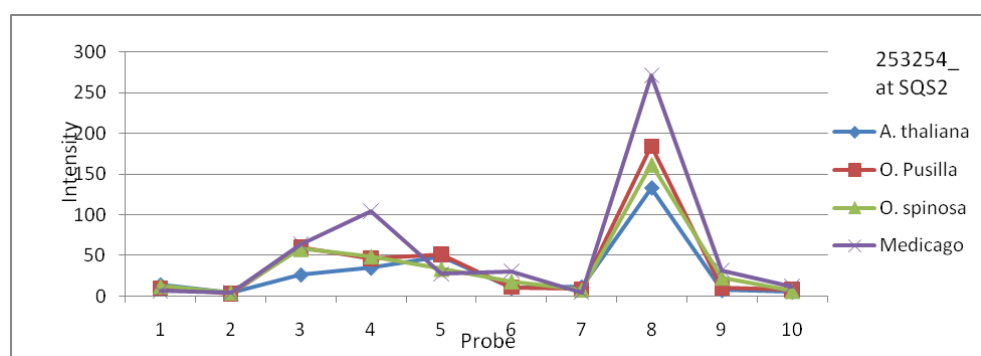
The ATH-1 GeneChip® array was designed primarily for hybridising *A. thaliana*. As such the signal value from PM oligonucleotides would be more highly conserved than between ATH-1 GeneChip® and gDNA from other plant species. One of the main assumptions of cross species transcriptomics is that the oligonucleotides cross-hybridise to the same genes within the cross-species. Thus the data for PM oligonucleotides was considered as the major target of interest for analysing hybridisation of individual oligonucleotides. The main aim was to identify where signal values of individual oligonucleotides (PM) was greater than that of the signal produced by gDNA from *A. thaliana*.

It was expected that *A. thaliana* would have a 100% signal for hybridisation against the PM oligonucleotide sequences. It was important to consider hybridisation values between PM pairs as hybridisation signal values from other species may indicate either a stronger hybridisation signal than *A. thaliana*.

#### 6.2.4 Squalene synthase hybridisation



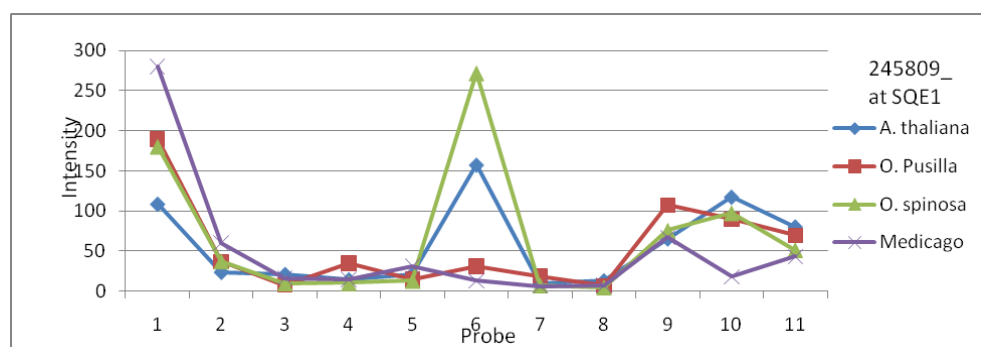
**Figure 6.5** Normalised signal value from gDNA hybridisation on ATH-1. 253206\_at, squalene synthase (At4g34640), (PM) hybridisation of four species compared; *A. thaliana*, *O. pusilla*, *O. spinosa* and *M. truncatula*.



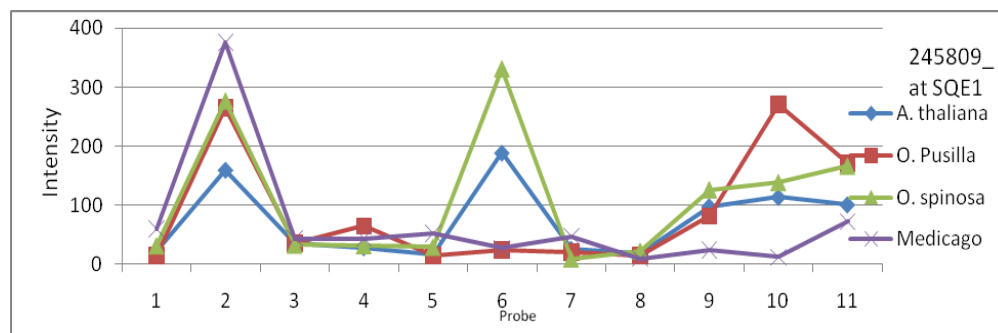
**Figure 6.6** Normalised signal value from gDNA hybridisation on ATH-1. 253254\_at, squalene synthase (At4g34650), (PM) hybridisation of four species compared; *A. thaliana*, *O. pusilla*, *O. spinosa* and *M. truncatula*.

A better match between the signal, would indicate a repeat sequence or the presence of multiple copies of oligonucleotides within the cross species genome. As a tool to reduce the amount of false hybridisation signals, GeneChip array masks were created based on signal values of individual probes. This approach was used as a way of analysing the hybridisation profiles of triterpene synthases.

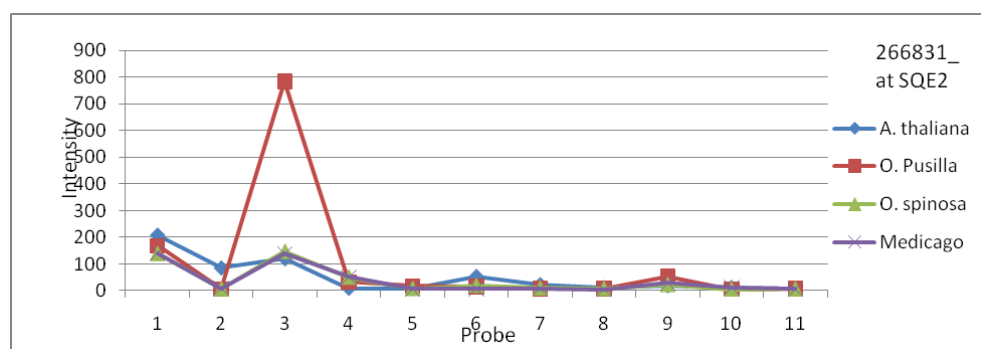
### 6.2.5 Squalene epoxidase hybridisation



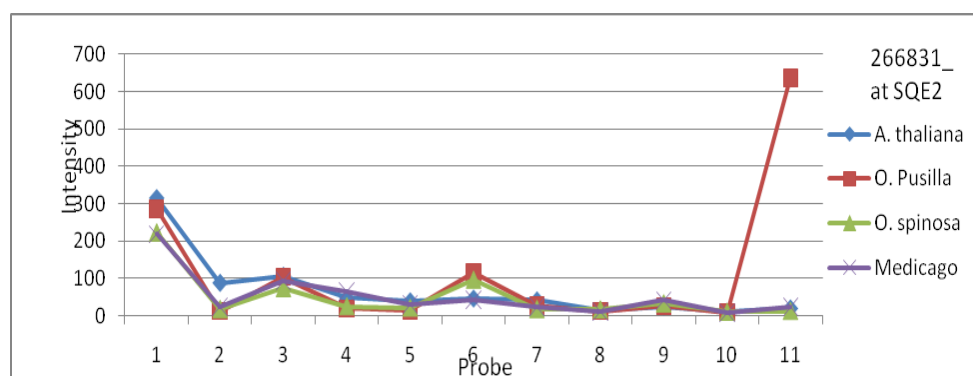
**Figure 6.7** Normalised signal value from gDNA hybridisation on ATH-1. 245809\_at, squalene epoxidase (At1g58440). (PM) hybridisation of four species compared; *A. thaliana*, *O. pusilla*, *O. spinosa* and *M. truncatula*.



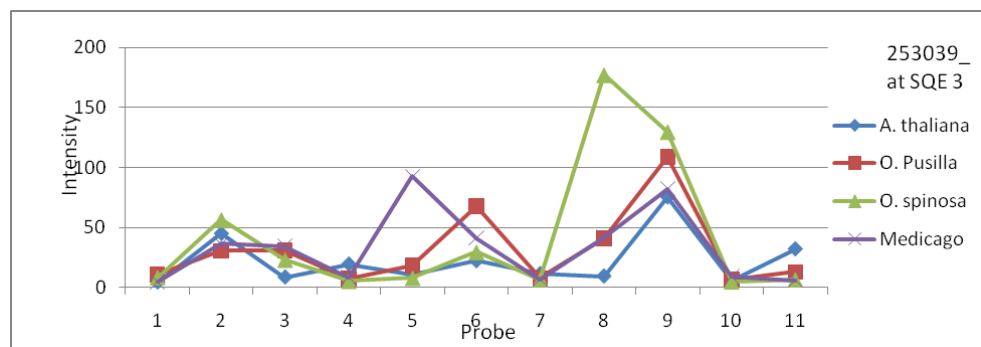
**Figure 6.8** Normalised signal value from gDNA hybridisation on ATH-1. 245809\_at, squalene epoxidase (At1g58440). (MM) hybridisation of four species compared; *A. thaliana*, *O. pusilla*, *O. spinosa* and *M. truncatula*.



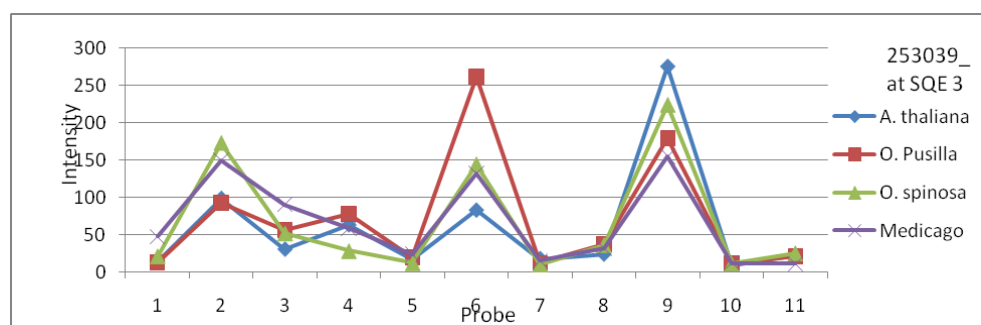
**Figure 6.9** Normalised signal value from gDNA hybridisation on ATH-1. 266831\_at, squalene epoxidase (At2g22830), (PM) hybridisation of four species compared; *A. thaliana*, *O. pusilla*, *O. spinosa* and *M. truncatula*.



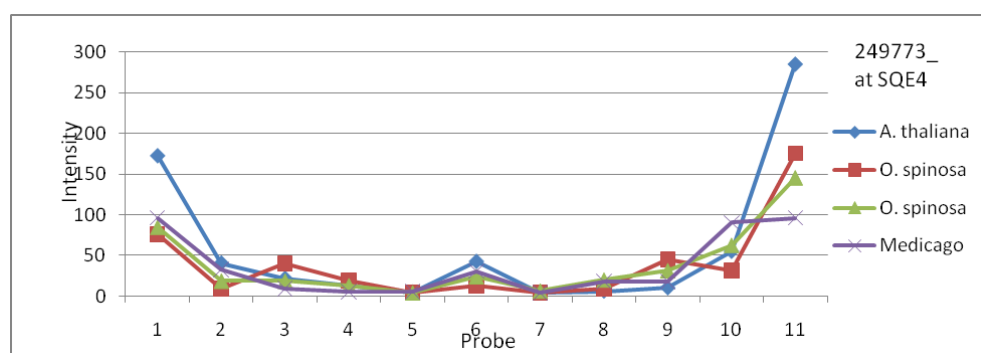
**Figure 6.10** Normalised signal value from gDNA hybridisation on ATH-1. 266831\_at, squalene epoxidase (At2g22830), (MM) hybridisation of four species compared; *A. thaliana*, *O. pusilla*, *O. spinosa* and *M. truncatula*.



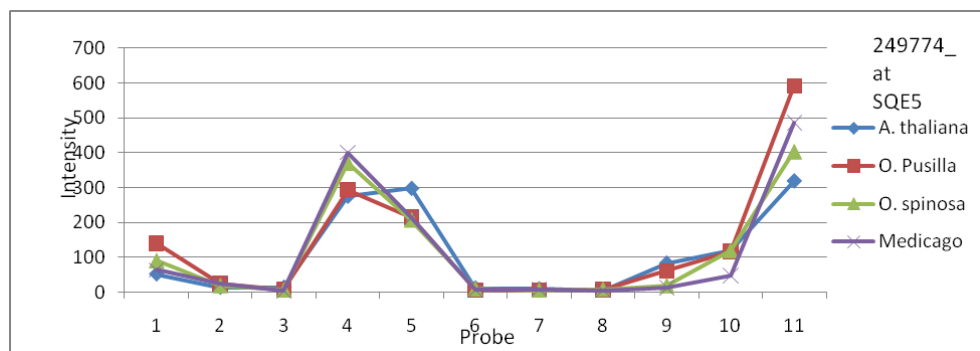
**Figure 6.11** Normalised signal value from gDNA hybridisation on ATH-1. 253039\_at, squalene epoxidase, (At4g37760), (PM) hybridisation of four species compared; *A. thaliana*, *O. pusilla*, *O. spinosa* and *M. truncatula*.



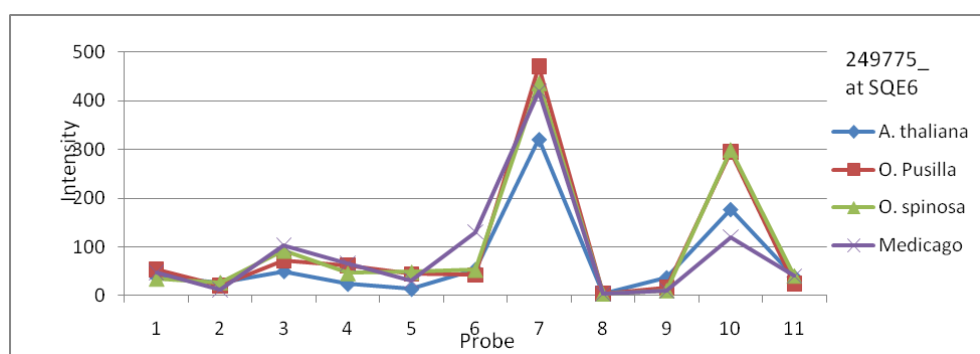
**Figure 6.12** Normalised signal value from gDNA hybridisation on ATH-1. 253039\_at, squalene epoxidase, (At4g37760), (MM) hybridisation of four species compared; *A. thaliana*, *O. pusilla*, *O. spinosa* and *M. truncatula*.



**Figure 6.13** Normalised signal value from gDNA hybridisation on ATH-1. 249773\_at, squalene epoxidase (At5g24140), (PM) hybridisation of four species compared; *A. thaliana*, *O. pusilla*, *O. spinosa* and *M. truncatula*.

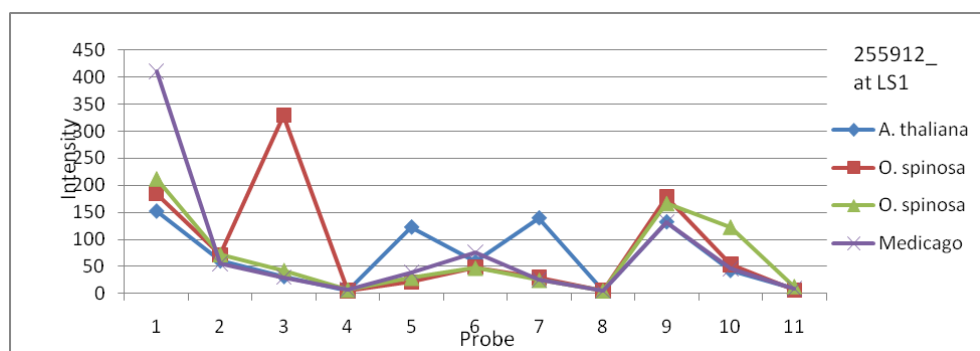


**Figure 6.14** Normalised signal value from gDNA hybridisation on ATH-1. 249774\_at, squalene epoxidase (At5g24150), (PM) hybridisation of four species compared; *A. thaliana*, *O. pusilla*, *O. spinosa* and *M. truncatula*.



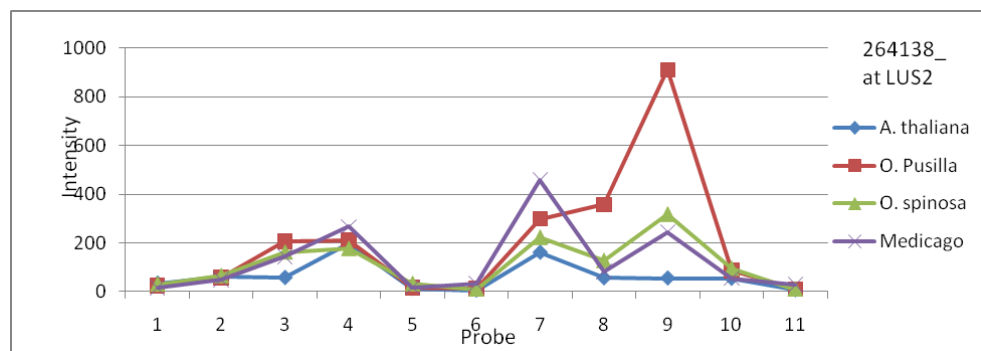
**Figure 6.15** Normalised signal value from gDNA hybridisation on ATH-1. 249775\_at, squalene epoxidase (At5g24160). (PM) hybridisation of four species compared; *A. thaliana*, *O. pusilla*, *O. spinosa* and *M. truncatula*.

## 6.2.6 Oxidosqualene cyclase hybridisation

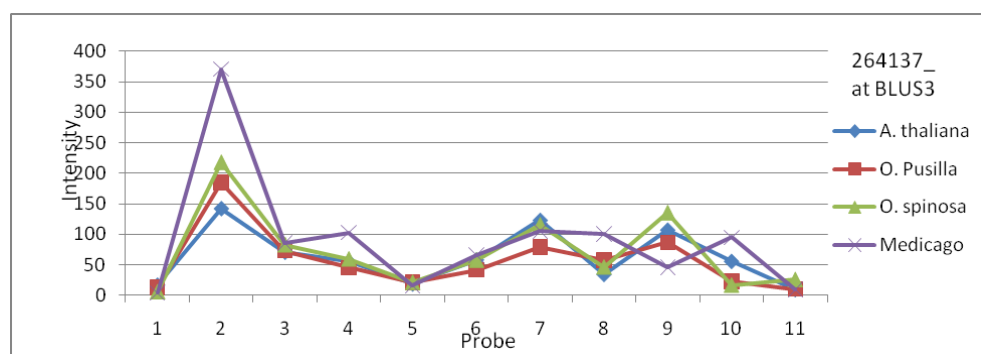


**Figure 6.16** Normalised signal value from gDNA hybridisation on ATH-1. 255912\_at, Lupeol synthase (At1g66960). (PM) hybridisation of four species compared; *A. thaliana*, *O. pusilla*, *O. spinosa* and *M. truncatula*.

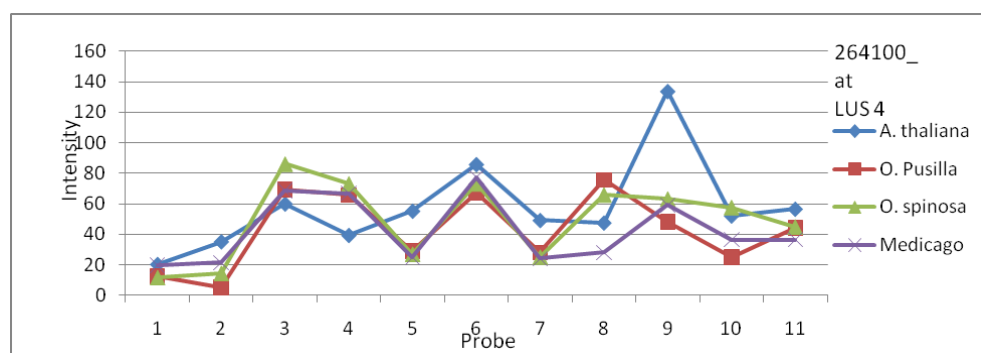




**Figure 6.17** Normalised signal value from gDNA hybridisation on ATH-1. 264138\_at,  $\beta$ -amyrin synthase (At1g78950). (PM) hybridisation of four species compared; *A. thaliana*, *O. pusilla*, *O. spinosa* and *M. truncatula*.

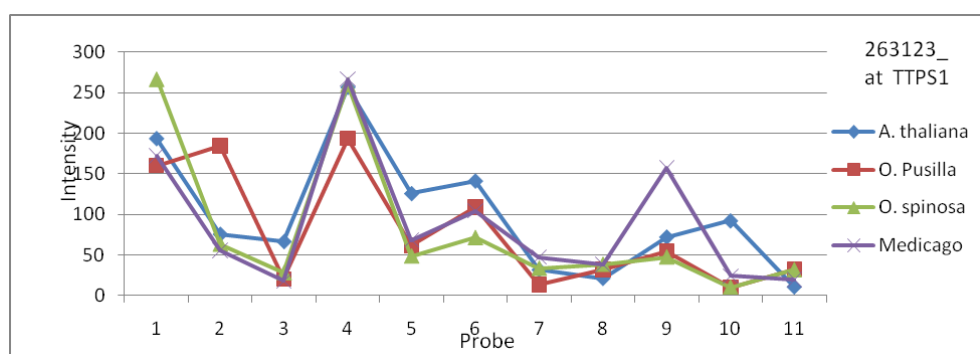


**Figure 6.18** Normalised signal value from gDNA hybridisation on ATH-1. 264137\_at,  $\beta$ -amyrin synthase (At1g78960). (PM) hybridisation of four species compared; *A. thaliana*, *O. pusilla*, *O. spinosa* and *M. truncatula*.

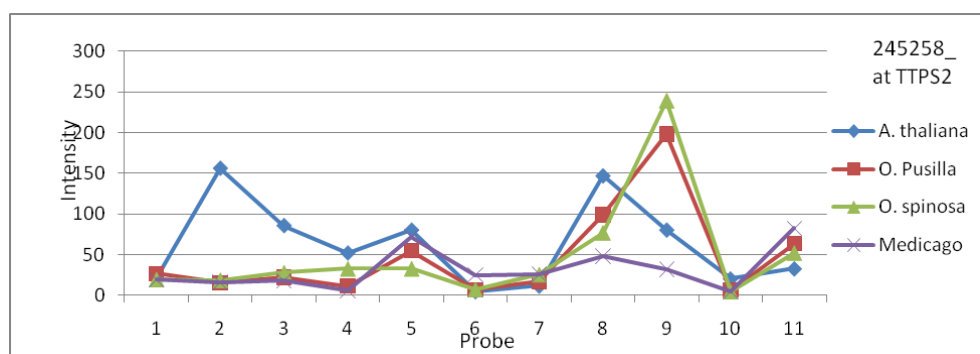


**Figure 6.19** Normalised signal value from gDNA hybridisation on ATH-1. 264100\_at, Lupeol synthase (At1g78970). (PM) hybridisation of four species compared; *A. thaliana*, *O. pusilla*, *O. spinosa* and *M. truncatula*.

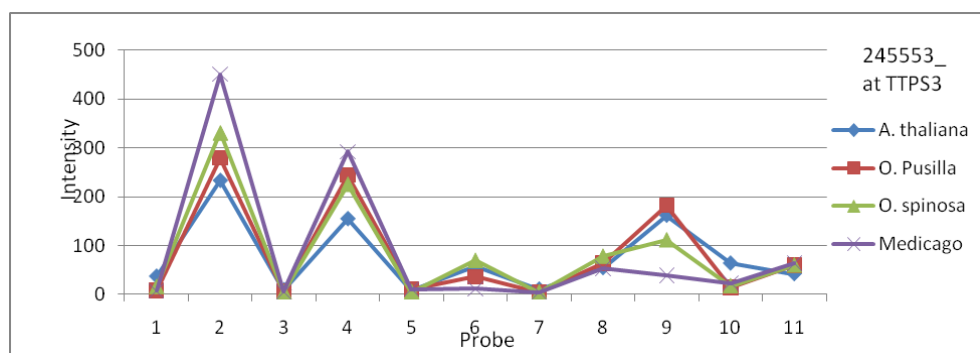
## 6.2.7 Putative triterpene synthase hybridisation



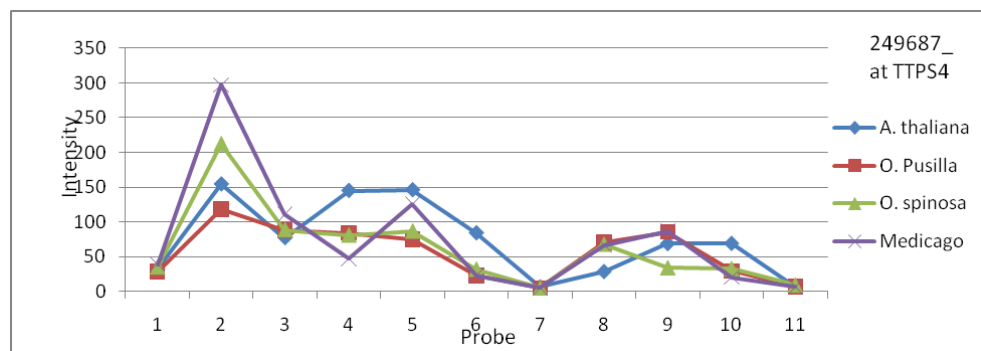
**Figure 6.20** Normalised signal value from gDNA hybridisation on ATH-1. 263123\_at, Triterpene synthase (At1g78500). (PM) hybridisation of four species compared; *A. thaliana*, *O. pusilla*, *O. spinosa* and *M. truncatula*.



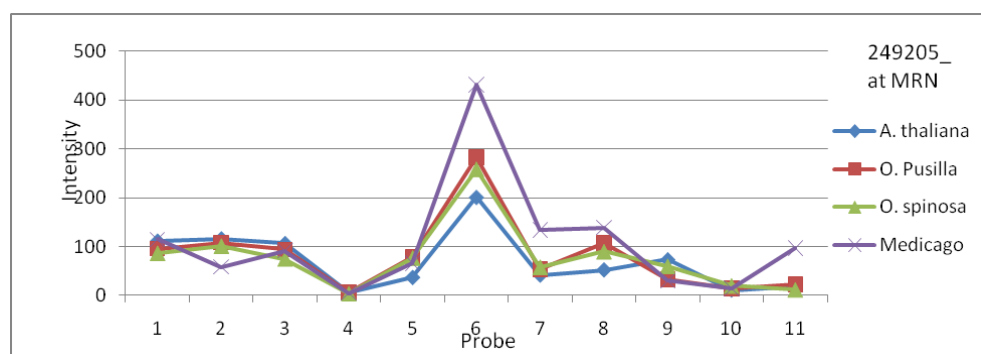
**Figure 6.21** Normalised signal value from gDNA hybridisation on ATH-1. 245258\_at, Triterpene synthase (At4g15340). (PM) hybridisation of four species compared; *A. thaliana*, *O. pusilla*, *O. spinosa* and *M. truncatula*.



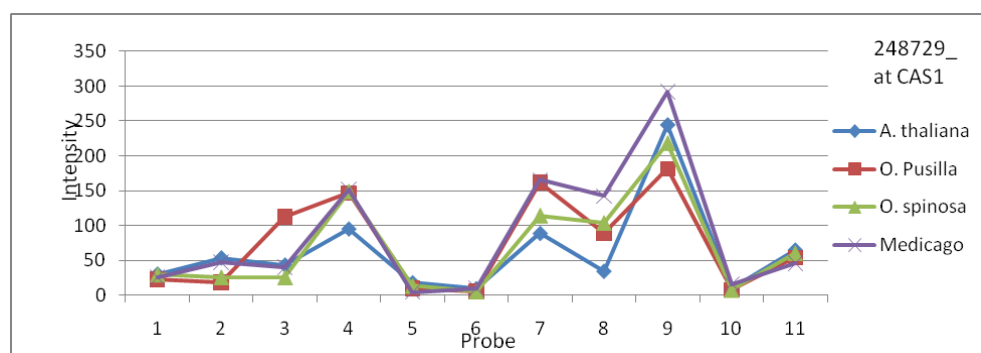
**Figure 6.22** Normalised signal value from gDNA hybridisation on ATH-1. 245553\_at, Triterpene synthase (At4g15370). (PM) hybridisation of four species compared; *A. thaliana*, *O. pusilla*, *O. spinosa* and *M. truncatula*.



**Figure 6.23** Normalised signal value from gDNA hybridisation on ATH-1. 249687\_at, Triterpene synthase (At5g36150). (PM) hybridisation of four species compared; *A. thaliana*, *O. pusilla*, *O. spinosa* and *M. truncatula*.



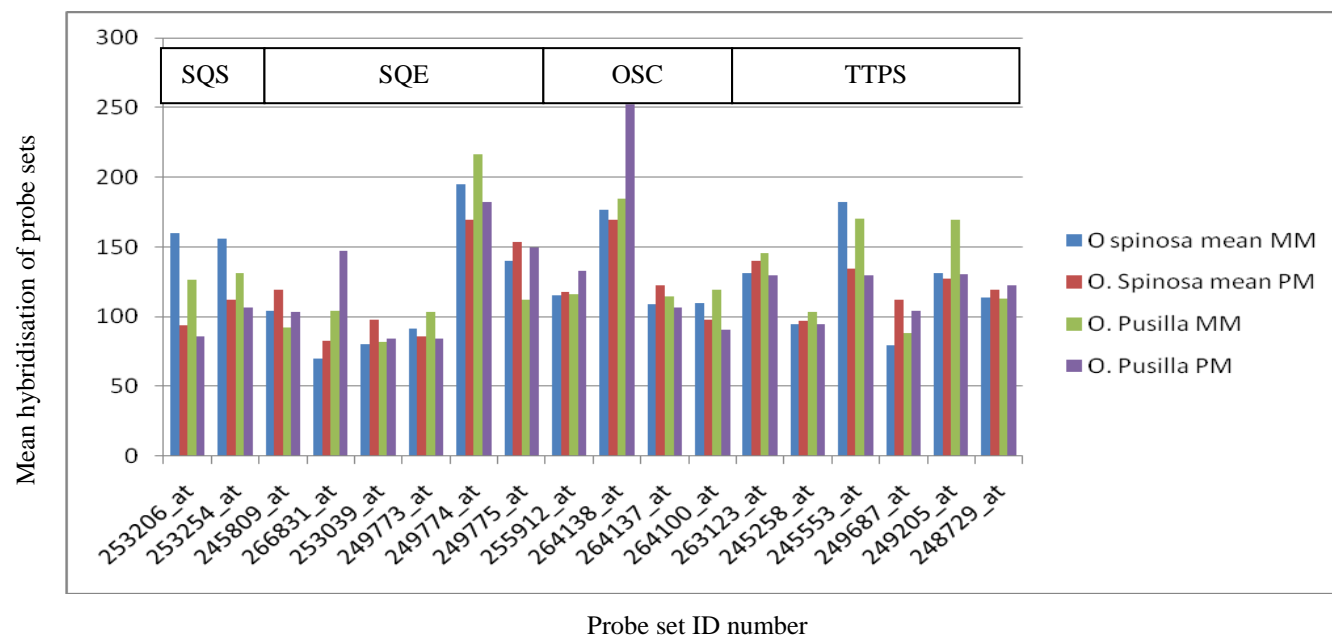
**Figure 6.24** Normalised signal value from gDNA hybridisation on ATH-1. 249205\_at, Triterpene synthase (At5g42600). (PM) hybridisation of four species compared; *A. thaliana*, *O. pusilla*, *O. spinosa* and *M. truncatula*.



**Figure 6.25** Normalised signal value from gDNA hybridisation on ATH-1. 248729\_at, Triterpene synthase (At5g48010). (PM) hybridisation of four species compared; *A. thaliana*, *O. pusilla*, *O. spinosa* and *M. truncatula*.

### 6.2.8 PM and MM analysis summary

The mean intensity value of PM and MM was summarised and plotted against each other in figure 6.27 in order to identify where gDNA hybridisation intensity differences occurred between species. The information was also used as a guideline for determining the mask threshold value. The mean hybridisation value was expected to be higher for PM than MM probes. For squalene synthase the difference in PM hybridisation between *O. spinosa* and *O. pusilla* was less than 10 (SQS), 253206\_at and 253254\_at. MM values for both species were higher than perfect match for both probe sets. For squalene epoxidase, Miss Match values were higher for 249773\_at and 249774\_at. 266831\_at (SQE2) the mean signal was higher than *O. spinosa* (PM, 75 and MM, 65) in *O. pusilla* for both PM (145) and MM (105). For 245809\_at, 253039\_at, 249775\_at the PM value mean PM value was higher than the MM value. Oxidosqualene cyclase 164138\_at, showed a higher PM mean hybridisation value in *O. pusilla* than in *O. spinosa*. Triterpene synthases 249205\_at, 245553\_at showed a higher MM value in both species than PM value.



**Figure 6.26** Normalised mean hybridisation signal for *O. spinosa* and *O. pusilla* PM and MM oligonucleotide probe sets. Genes listed in table 3.1; Squalene synthase (SQS), 253206\_at and 253254\_at. Squalene epoxidase (SQE), 245809\_at, 266831\_at, 253039\_at, 249773\_at, 249774\_at and 249775\_at. Oxidosqualene cyclase (OSC), 255912\_at, 264138\_at, 264137\_at and 264100\_at. Triterpene synthase (TTPS) 263123\_at, 245258\_at, 245553\_at, 249687\_at, 249205\_at and 248729\_at.

### 6.3 TRANSCRIPTIONAL PROFILING

Total RNA extracted from *O. pusilla* and *O. spinosa*, root and leaf was applied to ATH1-121501 GeneChip® arrays. Raw cell (.cdf) intensity files were loaded into GeneSpring (GX10, Silicon Genetics, CA, USA), using probe masks previously developed (pusilla 100 and spinosa 100). Each sample was treated as an individual experiment and was initially pre-normalised using Robust Multichip Average function (RMA) within GeneSpring. The aim was to evaluate the expression values between the main genes involved in the triterpene pathway.

#### 6.3.1 QC control figures and global expression difference between species

Analysis of triterpene synthase expression, RMA normalisation of transcripts and subsequent calculations was undertaken using GeneSpring (GX10), (Silicon Genetics, CA, USA). Initially a global investigation into the differentially expressed genes was undertaken using Gene view to outline differentially expressed genes. The aim of the global expression analysis was to investigate the differences between two mask files; pusilla 100 and spinosa100.

**Table 6.1** A general summary of Cdf mask file statistics. Cdf mask file PUSILLA 100 (threshold 100) was used to calculate expression values from *O. pusilla* and *O. spinosa* root and leaf RNA hybridised onto ATH-1.

<b>pusilla 100</b>	<b>[1 OP root]</b>	<b>[2 OP leaf]</b>	<b>[3 OS root]</b>	<b>[4 OS leaf]</b>
No. of Observations	22501.000	22501.000	22501.000	22501.000
No. of Missing Values	0.000	0.000	0.000	0.000
Minimum	-7.490	-2.775	-4.238	-2.205
Maximum	4.263	5.047	5.222	3.612
Mean	0.014	0.016	-0.003	0.015
Median	0.004	0.001	-0.010	0.003
Std. Deviation	0.277	0.266	0.303	0.275
No. Of Outliers	1784.000	1746.000	1975.000	1730.000

**Table 6.2** A summary of correlation between species and tissue types. Cdf mask file PUSILLA100 (threshold 100) was used to measure the correlation between *O. pusilla* and *O. spinosa* root and leaf RNA hybridised onto ATH-1.

Array Name	[1 OP root]	[2 OP leaf]	[3 OS root]	[4 OS leaf]
[1 OP root]	1.000	0.831	0.943	0.738
[2 OP leaf]	0.831	1.000	0.818	0.923
[3 OS root]	0.943	0.818	1.000	0.803
[4 OS leaf]	0.738	0.923	0.803	1.000

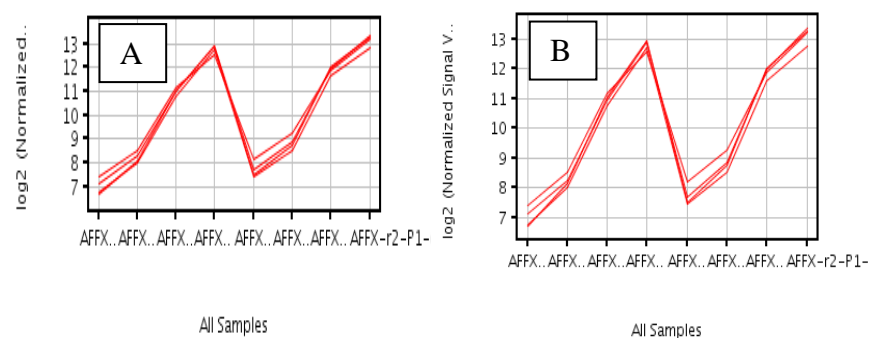
**Table 6.3** A general summary of statistics. Cdf mask file spinosa 100 (threshold 100) was used to calculate expression values from *O. pusilla* and *O. spinosa* root and leaf RNA hybridised onto ATH-1.

spinosa100	[1 OP root]	[2 OP leaf]	[3 OS root]	[4 OS leaf]
No. of Observations	22508.000	22508.000	22508.000	22508.000
No. of Missing Values	0.000	0.000	0.000	0.000
Minimum	-7.087	-2.594	-2.407	-4.041
Maximum	4.613	5.116	5.218	3.618
Mean	0.014	0.015	-0.005	0.014
Median	0.006	0.003	-0.013	0.002
Std. Deviation	0.275	0.265	0.306	0.273
No. Of Outliers	1855.000	1699.000	2102.000	1802.000

**Table 6.4** A summary of correlation between species and tissue types. Cdf mask file spinosa 100 (threshold 100) was used to assess correlation between *O. pusilla* and *O. spinosa* root and leaf RNA hybridised onto ATH-1.

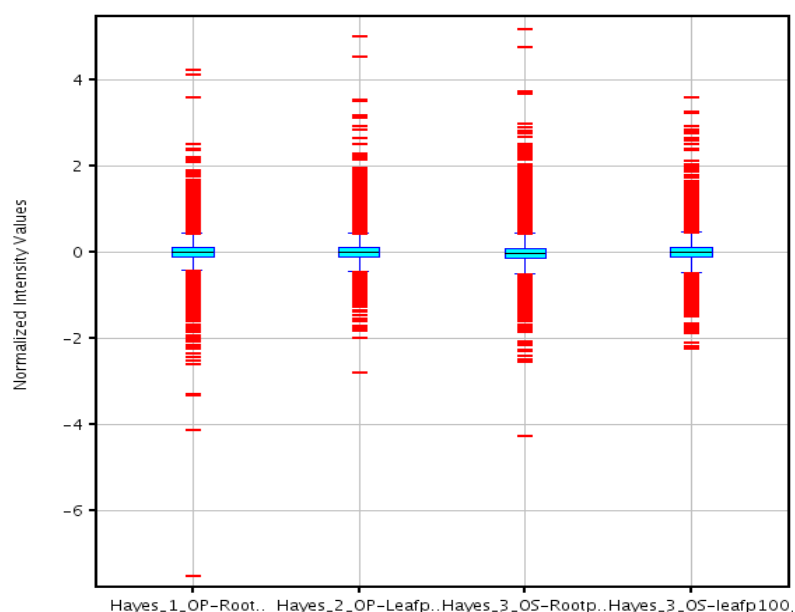
spinosa100	[1 OP root]	[2 OP leaf]	[3 OS root]	[4 OS leaf]
[1 OP root]	1.000	0.833	0.972	0.769
[2 OP leaf]	0.833	1.000	0.848	0.945
[3 OS root]	0.972	0.848	1.000	0.809
[4 OS leaf]	0.769	0.945	0.809	1.000

Quality control measures were used to assess the hybridisation efficiency of the RNA from *O. pusilla* and *O. spinosa*.



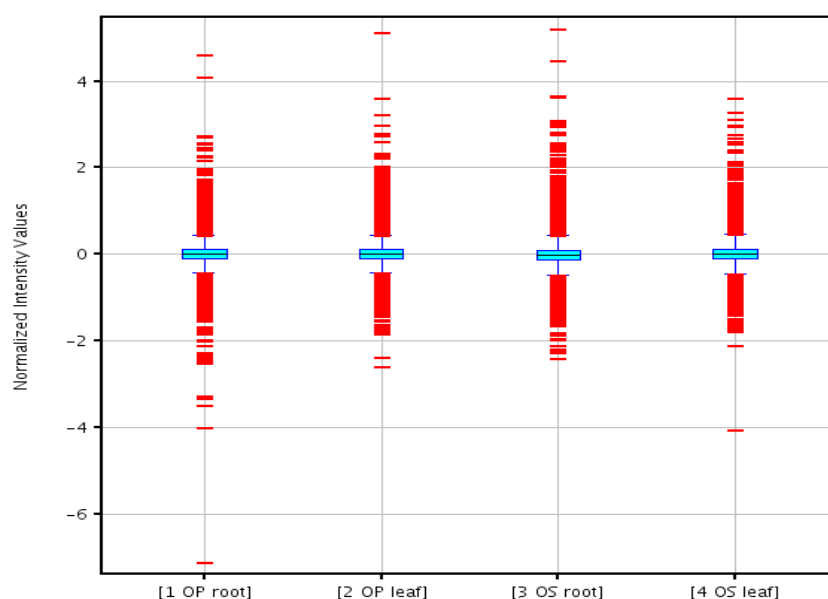
**Figure 6.27** Hybridisation controls for *O. pusilla* and *O. spinosa* root and leaf tissue, using mask files; A, pusilla 100 and B, spinosa 100.

A box whisper plot was used to display the overall normalised signal intensity values for *O. pusilla* and *O. spinosa* root and leaf using spinosa100 and pusilla100 mask files

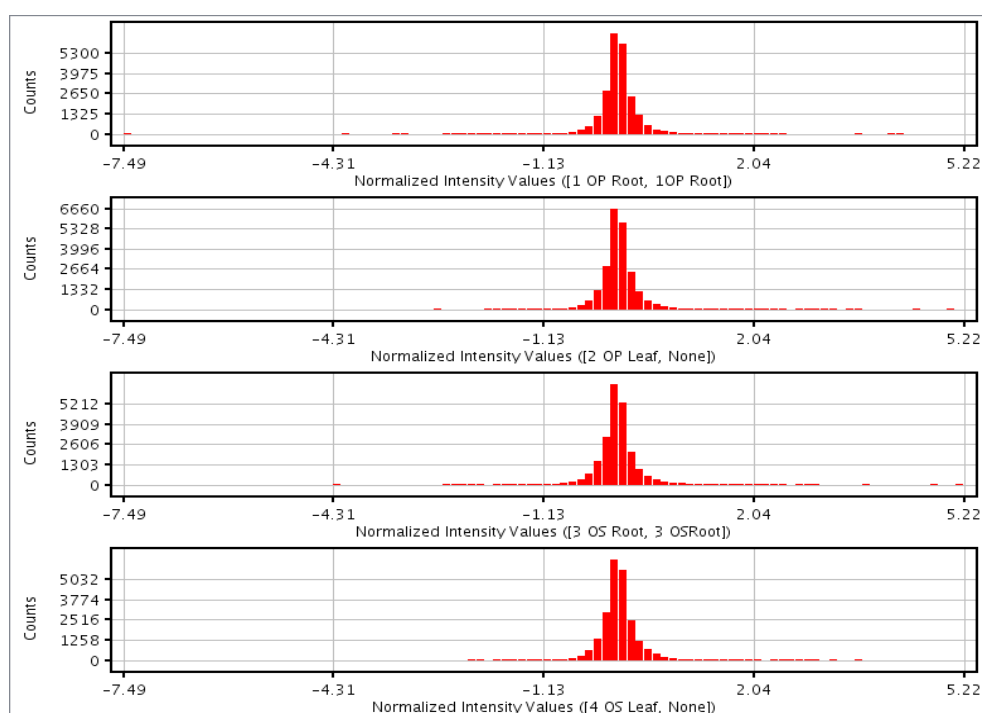


**Figure 6.28** Box whisper plot of the transcriptional profile using the mask pusilla100 (100 threshold), RMA normalised. Distribution intensity values of differentially expressed genes between *O. pusilla* root, *O. pusilla* leaf, *O. spinosa* root and *O. spinosa* leaf.

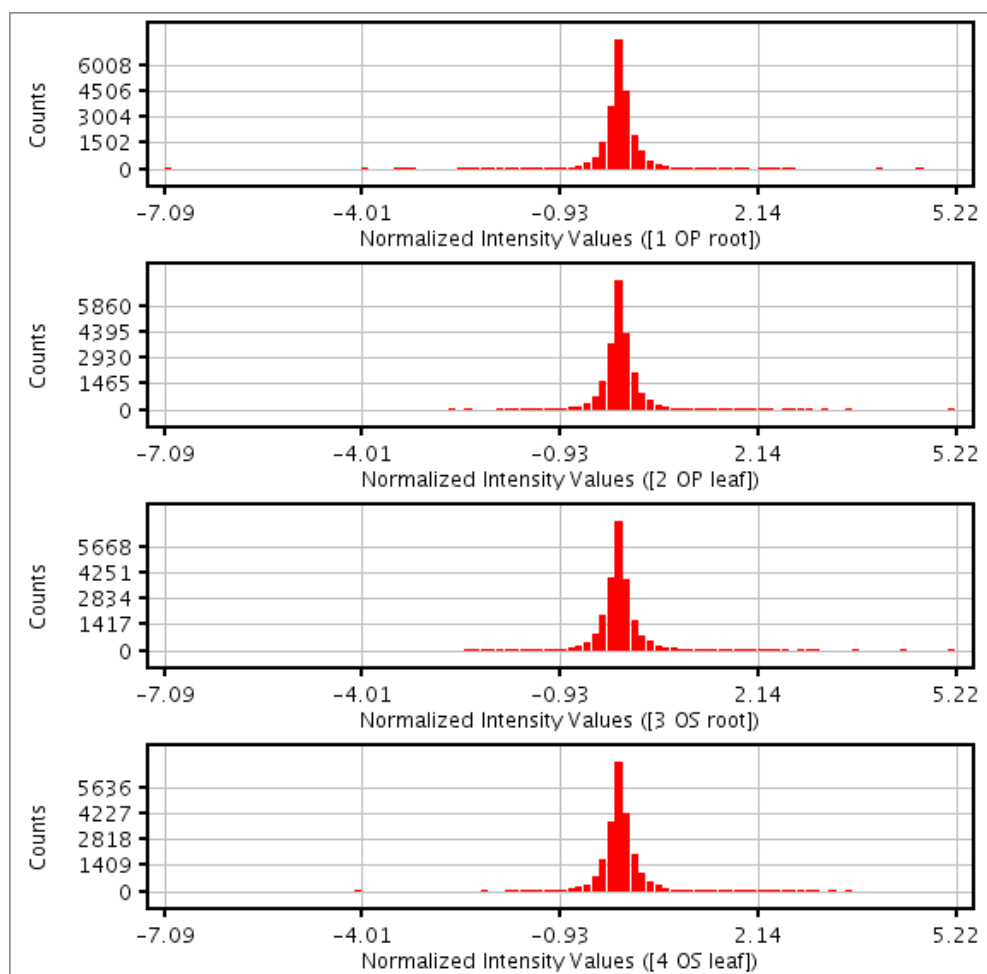




**Figure 6.29** Box whisker plot of the transcriptional profile using the mask *spinosa100* (100 threshold), RMA normalised. Distribution of intensity values from differentially expressed genes between *O. pusilla* root, *O. pusilla* leaf, *O. spinosa* root and *O. spinosa* leaf.



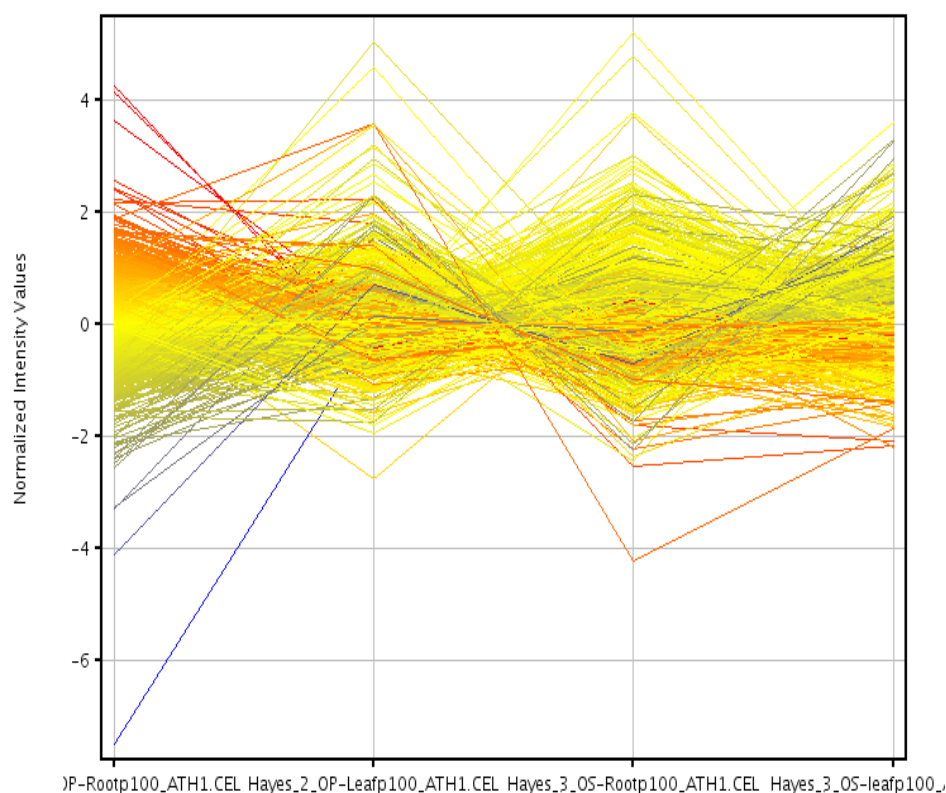
**Figure 6.30** A Histogram of the transcriptional profile using the mask *pusilla100* (100 threshold), RMA normalised, showing normal distribution of signal intensities between *O. pusilla* root, *O. pusilla* leaf, *O. spinosa* root and *O. spinosa* leaf.



**Figure 6.31** A Histogram of the transcriptional profile using the mask spinosa100 (100 threshold), RMA normalised, showing normal distribution of signal intensities between *O. pusilla* root, *O. pusilla* leaf, *O. spinosa* root and *O. spinosa* leaf.

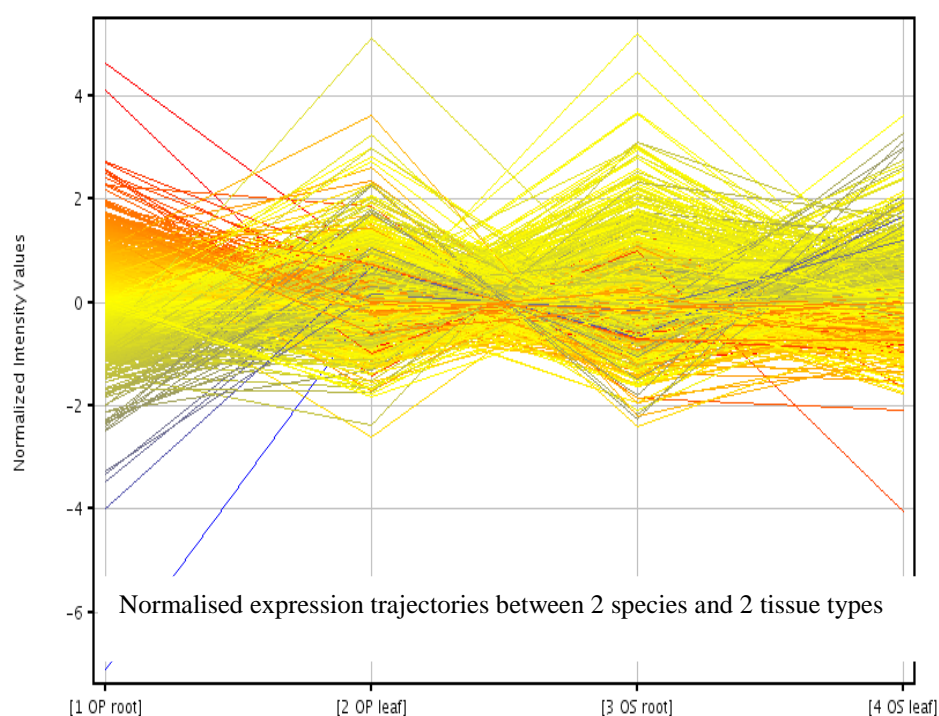
#### 6.4 DIFFERENTIALLY EXPRESSED GENES

Global maps of differentially expressed genes were generated between *O. pusilla* root, *O. pusilla* leaf, *O. spinosa* root and *O. spinosa* leaf using both mask files. The interaction charts in figures 6.33, 6.34, 6.35 and 6.36 were used to select both up-regulated and down regulated genes between the two plant species and tissue types. The interaction charts allowed 20544 probe sets from spinosa100 and 20567 probe sets *O. pusilla*, pusilla100 to be analysed. This analysis focussed on 20 probe sets involved in triterpene biosynthesis.

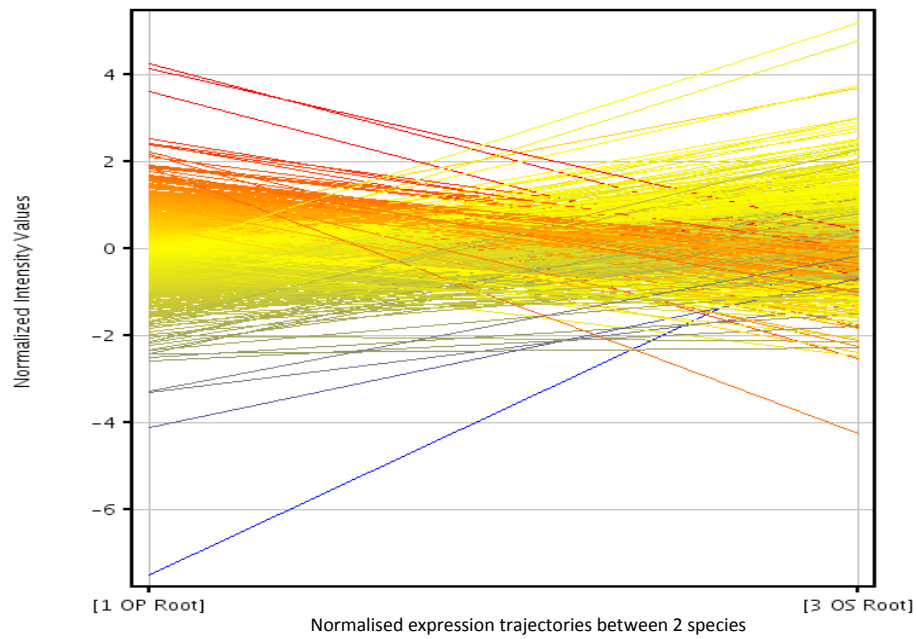


Normalised expression trajectories between 2 species and 2 tissue types

**Figure 6.32** An overview of normalised expression trajectories of differentially expressed genes between *O. pusilla* root (OP-Rootp100), *O. pusilla* leaf (OP-Leafp100), *O. spinosa* root (OS-Root p100) and *O. spinosa* leaf (OS-leafp100). Signal intensities from 20544 differentially expressed probe sets using the CDF masking file (PUSILLA100) threshold of 100 from hybridised *O. pusilla* gDNA on ATH1-121501 GeneChip® array. The colour red indicates up regulation and blue indicates down regulation of the genes, yellow depicts stable expression levels.

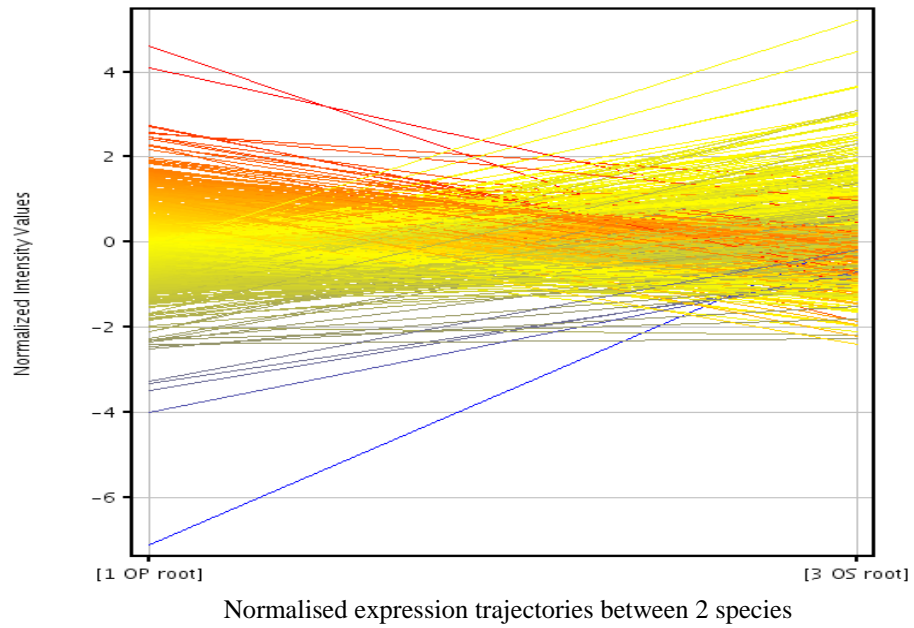


**Figure 6.33** An overview of normalised expression trajectories of differentially expressed genes between *O. pusilla* root (OP Root), *O. pusilla* leaf (OP Leaf), *O. spinosa* root (OS Root) and *O. spinosa* leaf (OS leaf). Signal intensities from 20567 differentially expressed probe sets using the CDF masking file (SPINOSA100) threshold of 100 from hybridised *O. spinosa* gDNA on ATH1-121501 GeneChip® arrays. The colour red indicates up regulation and blue indicates down regulation of the genes, yellow depicts stable expression levels.



Differentially expressed root samples

**Figure 6.34** An overview of normalised expression trajectories of differentially expressed genes between *O. pusilla* root (OP Root), and *O. spinosa* root (OS Root). Signal intensities from 20544 differentially expressed probe sets using the CDF masking file (PUSILLA100) threshold of 100 from hybridised *O. pusilla* gDNA on ATH1-121501 GeneChip® array. The colour red indicates up regulation and blue indicates down regulation of the genes, yellow depicts stable expression levels.



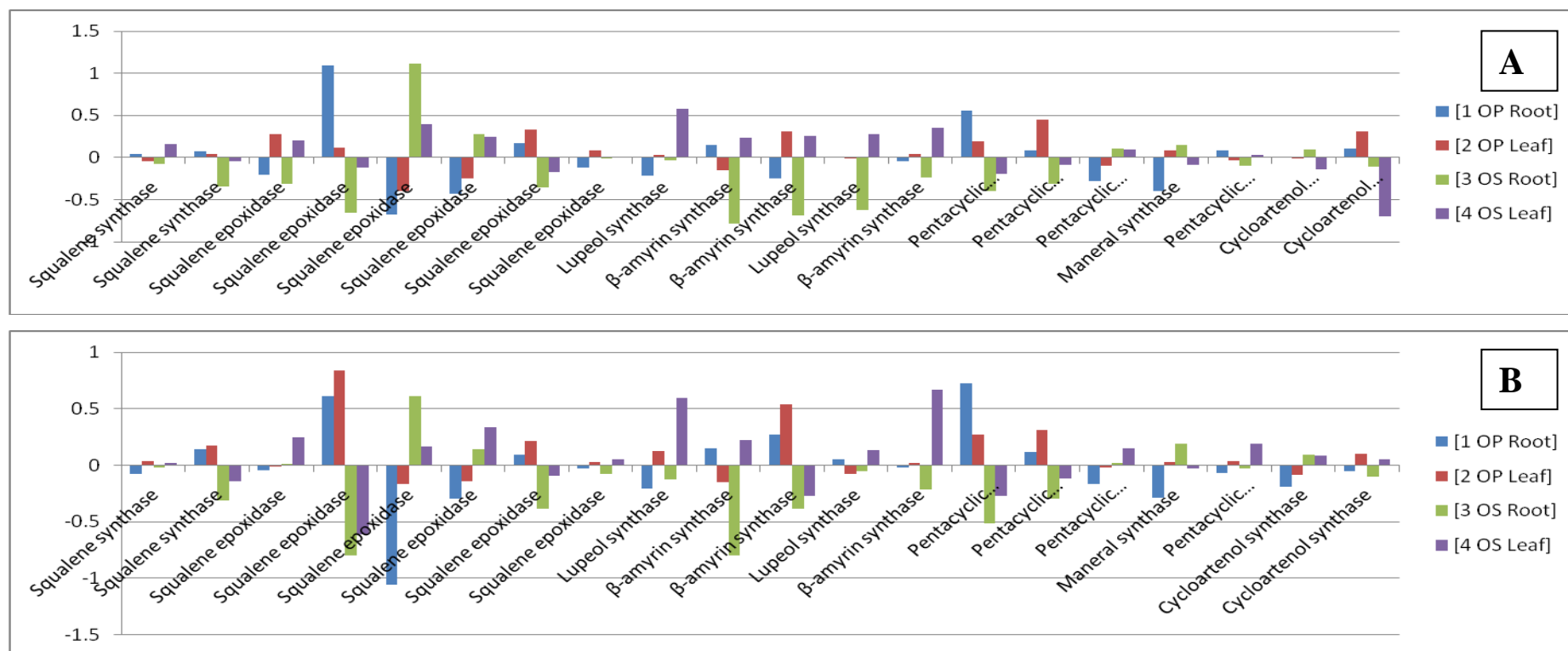
**Figure 6.35** An overview of normalised expression trajectories of differentially expressed genes between *O. pusilla* root (OP Root), and *O. spinosa* root (OS Root). Signal intensities from 20544 differentially expressed probe sets CDF masking file (spinosa100), threshold of 100 from hybridised *O. spinosa* gDNA on ATH1-121501 GeneChip® array. The colour red indicates up regulation and blue indicates down regulation of the genes, yellow depicts stable expression levels.

## **6.5 GENE EXPRESSION WITHIN THE TRITERPENE PATHWAY**

### **6.5.1 Gene expression differences between two species of Restharrow *O. pusilla* and *O. spinosa***

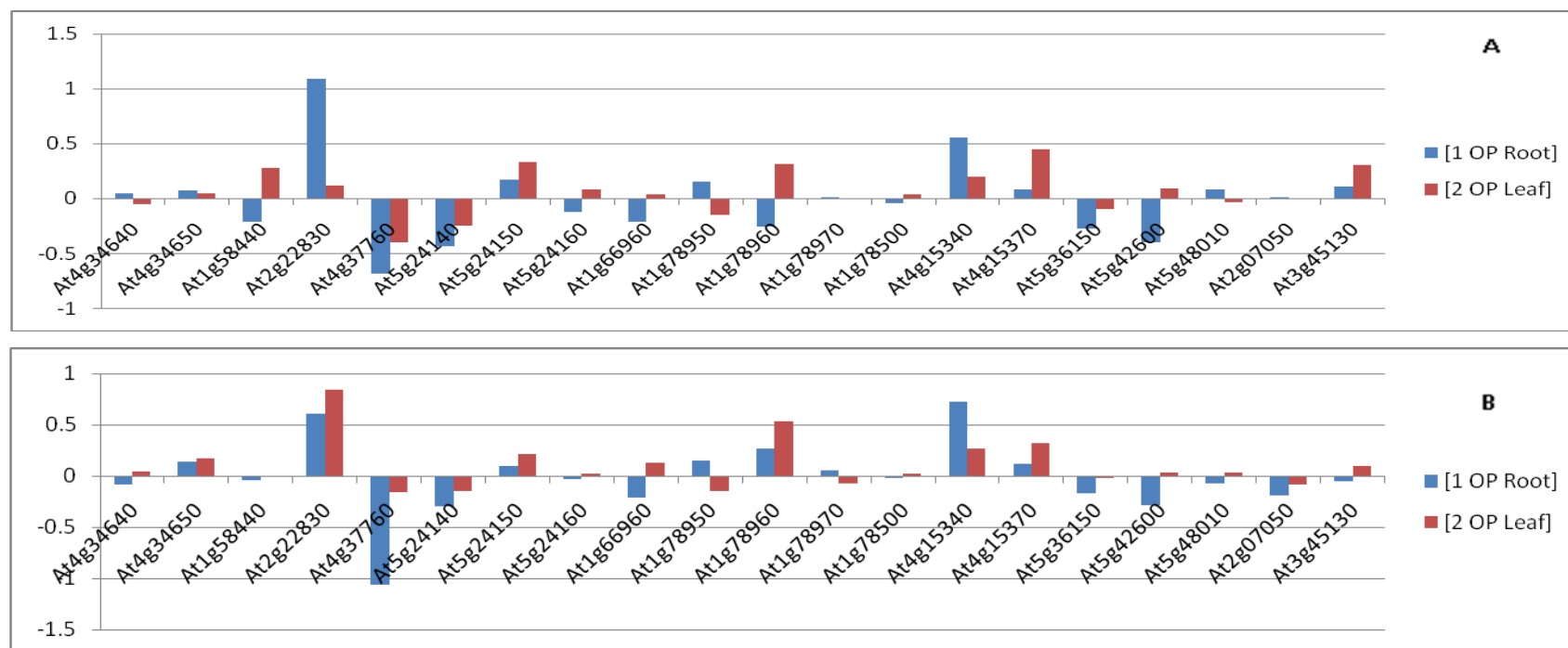
The gene set as outlined in table 3.1, was used to analyse gene expression values across the triterpene biosynthetic pathway. The transcriptional profile was aligned with the biochemical pathway.

The individual expression values across the two separate hybridisation masks, were analysed as a comparison between data sets. The results showed a small amount of variation which was expected. The same pattern occurred between tissue types across both cdf. mask files for most of the genes involved within the pathway. This was a good indication that cross hybridisation within the individual oligonucleotides had little effect on the experiment. The means of expression values from both masks pusilla100 and spinosa100 were calculated. *O. pusilla* and *O. spinosa* are closely related Restharrow species. As the hybridisation values within the individual oligonucleotides were closely matched along with the expression it was a reasonable approach to utilise the data from the two masks as duplicates.

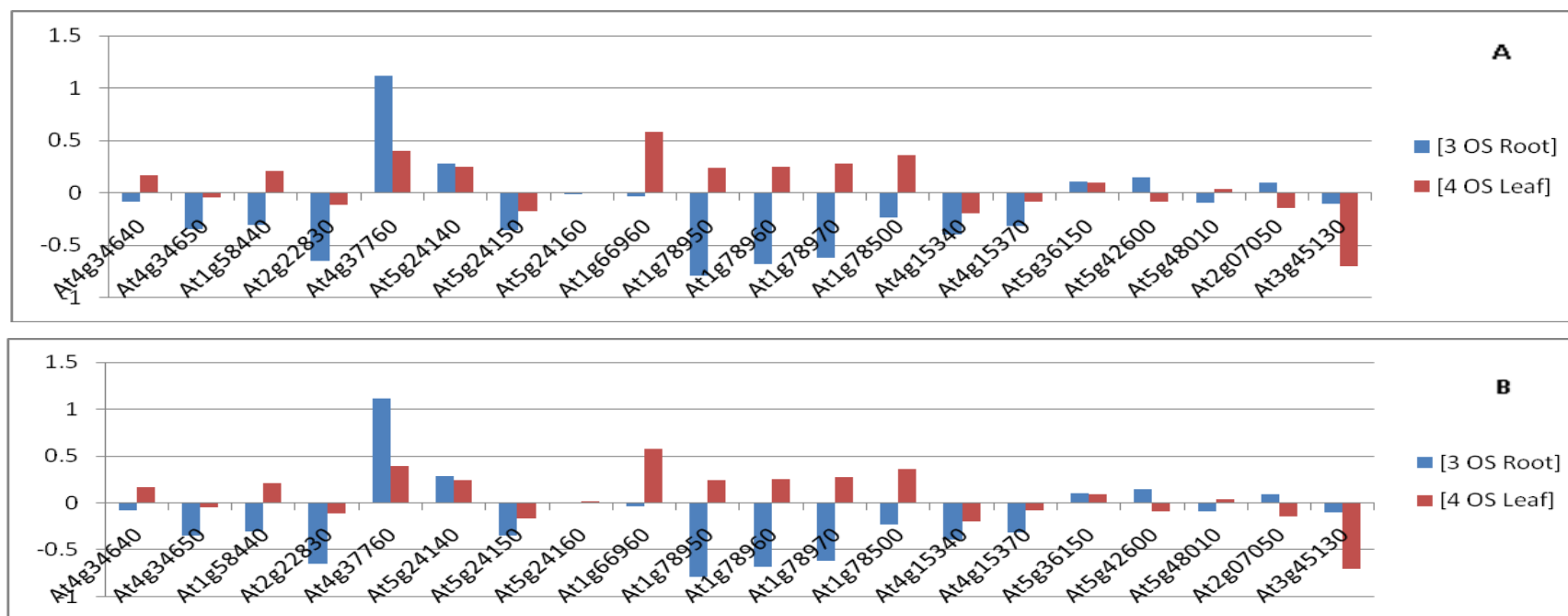


**Figure 6.36** Overview of normalised signal expression value of genes involved in the triterpene biosynthetic pathway.. Transcriptional profile between *O. pusilla* Root, *O. pusilla* leaf, *O. spinosa* root and *O. spinosa* leaf. A) CDF masking file (PUSILLA100) threshold of 100 from hybridised *O. pusilla* gDNA on ATH1. B) CDF masking file (spinosa100) threshold of 100 from hybridised *O. spinosa* gDNA on ATH1. AGI codes, probe set ID and gene nomenclature are present in table 6.1 for interpretation.



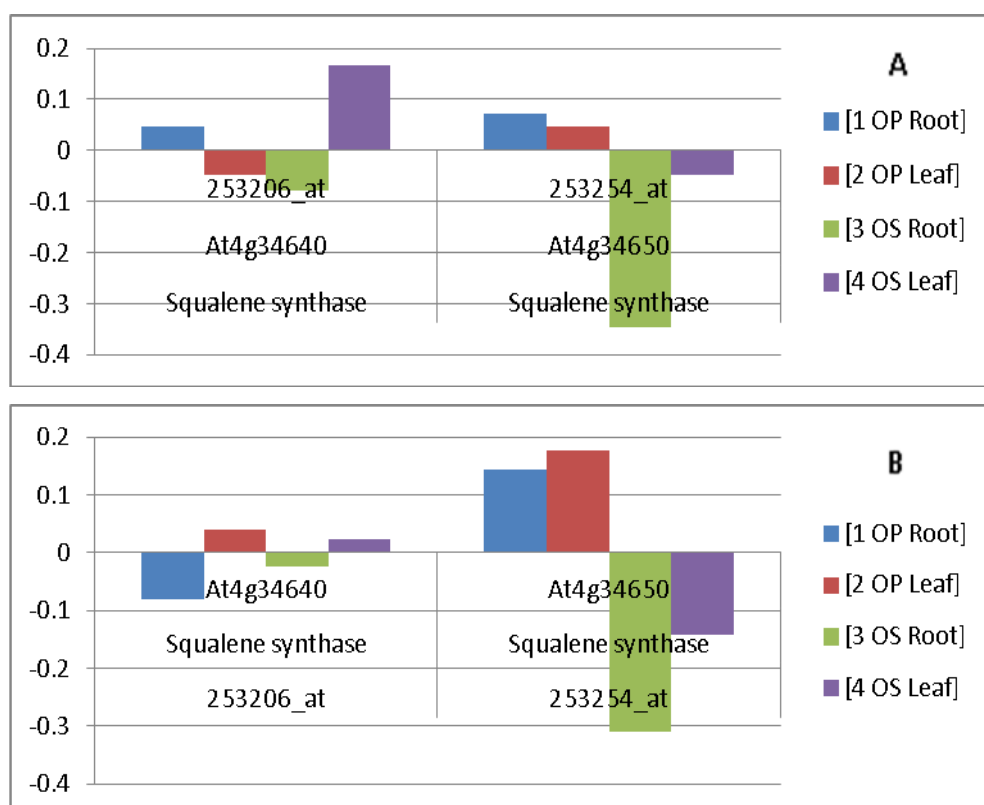


**Figure 6.37** Normalised signal expression value of differentially expressed oxidosqualene cyclases involved in the triterpene biosynthetic pathway. Transcriptional profile between *O. pusilla* Root, *O. pusilla* leaf, *O. spinosa* root and *O. spinosa* leaf. The mean signal value between was plotted for; A) CDF masking file (PUSILLA100) threshold of 100 from hybridised *O. pusilla* gDNA on ATH1. B) CDF masking file (spinosa100) threshold of 100 from hybridised *O. spinosa* gDNA on ATH1. AGI codes, probe set ID and gene nomenclature are present in table 3.1 for interpretation.



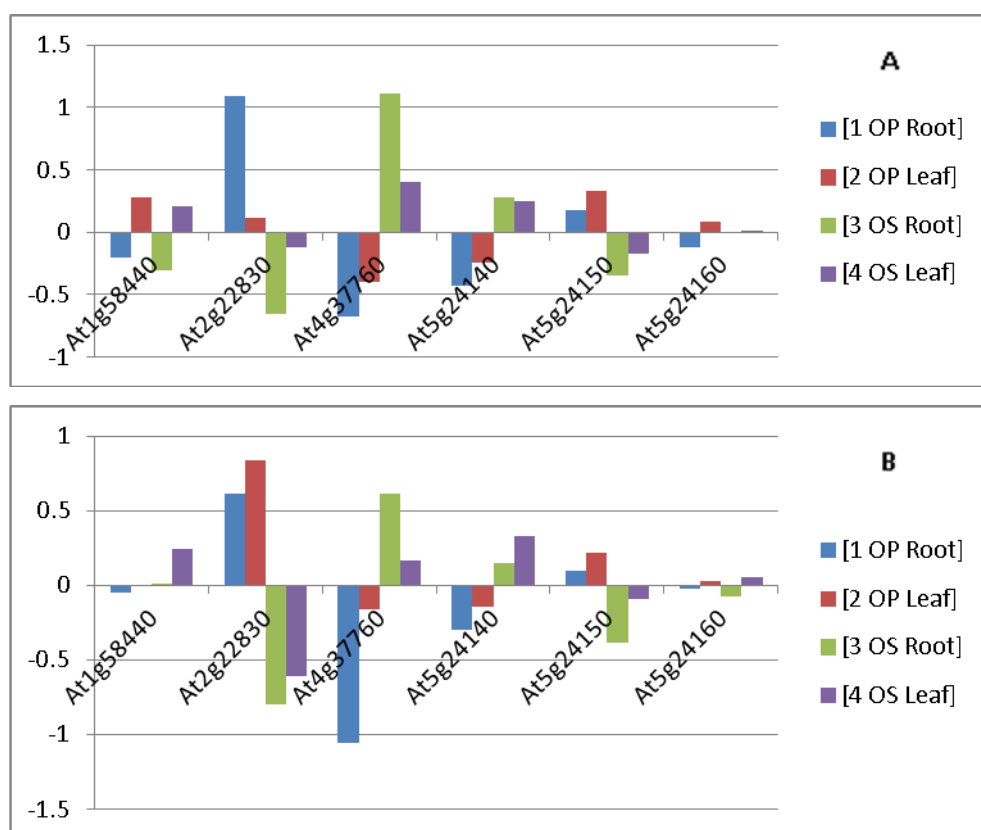
**Figure 6.38** Normalised signal expression value of differentially expressed. Oxidosqualene cyclases involved in the triterpene biosynthetic pathway. Transcriptional profile between *O. pusilla* Root, *O. pusilla* leaf, *O. spinosa* root and *O. spinosa* leaf. The mean signal value between was plotted for; A) CDF masking file (PUSILLA100) threshold of 100 from hybridised *O. pusilla* gDNA on ATH1. B) CDF masking file (spinosa100) threshold of 100 from hybridised *O. spinosa* gDNA on ATH1. AGI codes, probe set ID and gene nomenclature are present in table 3.1 for interpretation.

## 6.5.2 Gene expression squalene synthase



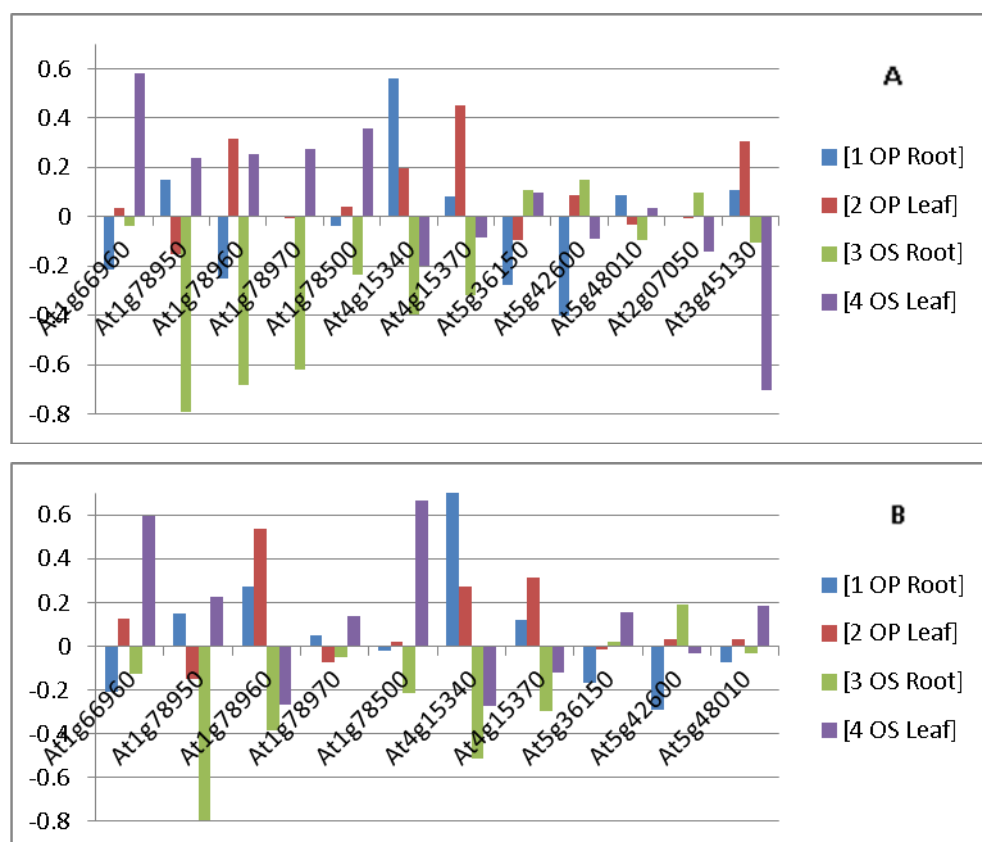
**Figure 6.39** Normalised signal expression value of differentially expressed squalene synthases involved in the triterpene biosynthetic pathway. Transcriptional profile between *O. pusilla* root, *O. pusilla* leaf, *O. spinosa* root and *O. spinosa* leaf. The mean signal value between was plotted for; A) CDF masking file (PUSILLA100) threshold of 100 from hybridised *O. pusilla* gDNA on ATH1. B) CDF masking file (spinosa100) threshold of 100 from hybridised *O. spinosa* gDNA on ATH1. AGI codes, probe set ID and gene nomenclature are present in table 3.1 for interpretation.

### 6.5.3 Gene expression squalene epoxidase



**Figure 6.40** Normalised signal expression value of differentially expressed squalene epoxidases involved in the triterpene biosynthetic pathway. Transcriptional profile between *O. pusilla* root, *O. pusilla* leaf, *O. spinosa* root and *O. spinosa* leaf. The signal value was plotted for; A) CDF masking file (PUSILLA100) threshold of 100 from hybridised *O. pusilla* gDNA on ATH-1. B) CDF masking file (spinosa100) threshold of 100 from hybridised *O. spinosa* gDNA on ATH1. AGI codes, gene nomenclature are presented in table 3.1 for interpretation.

### 6.5.4 Gene expression oxidosqualene cyclase



**Figure 6.41** Normalised signal expression value of differentially expressed oxidosqualene cyclases involved in the triterpene biosynthetic pathway. Transcriptional profile between *O. pusilla* Root, *O. pusilla* leaf, *O. spinosa* root and *O. spinosa* leaf. The mean signal value between was plotted for; A) CDF masking file (PUSILLA100) threshold of 100 from hybridised *O. pusilla* gDNA on ATH1. B) CDF masking file (spinosa100) threshold of 100 from hybridised *O. spinosa* gDNA on ATH1. AGI codes, probe set ID and gene nomenclature are present in table 3.1 for interpretation.

## 6.6 DISCUSSION

Transcriptional profiling using cross species hybridisation an Affymetrix GeneChip arrays is an accepted method of transcriptional profiling of organisms where very little genomic sequence data is available. Triterpene

synthases for *O. spinosa* and *O. pusilla* root and leaf RNA were transcriptionally profiled on ATH-1 GeneChip (Affymetrix, Santa Clara, USA). *O. spinosa* gDNA hybridisation signal value was normalised using RMA analysis using cross species hybridisation discovery the Software PIGEON (Unpublished Mayes and Lai (2010)). The number of probe pairs against probes sets lost at different thresholds 20-1000 intensity values was plotted. For *O. spinosa* gDNA, threshold 100 was used to generate a masking file, also using PIGEONS. At this threshold the number of probe sets retained was 20544 and probe pairs was 55543. Thus 91% of probes were retained and 23% of probe pairs were retained. This was considered an acceptable level of probe pair loss in comparison to the number of probes retained. In *O. pusilla* the number of probe sets retained was 20567 (90%) and probe pairs retained was 56500 (23%) at signal threshold of 100.

As deduced from the triterpene pathway information, genome information and the hypotheses developed in chapter 1, the genes that are most likely to be involved to affect  $\alpha$ -onocerin biosynthesis are Squalene epoxidases based on Rowans hypothesis (1971) and hypothesis 2. Hypothesis 1 would suggest there should be some key differences between the Oxidosqualene cyclases either at the sequence level or transcriptional profile. The primary focus of the cross species transcriptional profiling was to deduce whether the ATH-1 GenChip (Affymetrix, Santa Clara, USA) could be used as a platform for transcriptional profiling in *Restharrow*. The secondary focus was to determine whether there was any differences within the transcriptional profiles between two closely related species (multi-species cross species transcriptomics), and different tissue types.

The gDNA hybridisation threshold data was initially analysed. Genomic DNA from *A. thaliana* should theoretically have the highest signal values using the PM oligonucleotides. This was not the case across 21 genes involved in triterpene biosynthesis. The probe set signal values from *M. truncatula*, *O. pusilla* and *O. spinosa*, followed the same pattern as the *A. thaliana* gDNA. Differences in signal oligonucleotide probe values

between each species may be due to differences in hybridisation efficiencies as a result of amino acid sequence variation.

The gDNA PM signal values were used to identify sequence variation between *O. pusilla* and *O. spinosa*. Within the Squalene epoxidase SQE1 oligonucleotide 6 (SQE1-6), (known to cyclise 2,3;22,23-dioxidosqualene) there were differences between *M. truncatula* and *O. pusilla*. The highest intensity was 272 in *O. spinosa* and 157 in *A. thaliana*, whereas in *M. truncatula* and *O. pusilla* the intensity was 13 and 30 respectively. This indicates that there was more sequence variation between *O. pusilla* and *A. thaliana* than there was between *A. thaliana* and *O. spinosa*. The intensity values for the other 10 oligonucleotides for SQE1 matched closely). This may mean functional differences may occur between squalene epoxidase in *O. pusilla* and *O. spinosa*. Within squalene SQE2-3, *O. pusilla* had a signal value 784, although it may represent a difference in sequence, a high signal value represents cross hybridisation onto a different oligonucleotide. As a result SQE2-3 was be disregarded and removed from the expression analysis using the mask file.

Within SQE3 the highest intensity was 180, suggesting the sequence differences occurred for the oligonucleotides rather than it being a cross hybridisation event onto a different oligonucleotide. Primarily the shybridisation intensity for SQE3-5 in *M. truncatula* was 93, whereas *A. thaliana* was 8; this may equate to a sequence difference. SQE3-8 gave a signal value of 177 in *O. spinosa* in contrast *O. pusilla* had an intensity of 40 and *A. thaliana* had a signal intensity of 9. Oligonucleotide SQE3-8 may indicate that sequence variation between *O. pusilla* and *O. spinosa* may occur.

The oligonucleotide sequences where high intensity values occurred were;

SQE1-6 (GAACCCTCGACCATTAACACTTGTC)  
SQE2-3 (AGGCCAATTCGCAACCTTAATGACA)  
SQE3-8 (CATCAGGACCTGTGGCTTTGCTATC)

These sequences were translated into amino acid sequences and aligned using clustalW (EBI.ac.uk)

Homolog	Amino acid alignment
SQE2	GGGMTVALADIVVLRDLLRPIRNLNKEALSKYIESFYTLRKPVASTINTLADALYKVFL 478
SQE3	GGGMTVALSDIVILRDLLNPLVDLTNKEALSKYIESFYTLRKPVASTINTLAGALYKVFL 410
SQE1	GGGMTVALADIVVLRNLLRPLRDLSDGASLCKYLESFYTLRKPVAAATINTLANALYQVFC 416
	*****:***:***:***:***: :*: :*: :*: :*: :*: :*: :*: :*: :*: :*
SQE2	ASSDEARTEMREACFDYLSLGGVFSSGPGVALLSGLNPRPLSLVLHFFAVAIYAVCRMLLP 538
SQE3	ASPDDARSEMRRACFDYLSLGGVCSGPGVALLSGLNPRPMSLVLHFFAVAIYAVGRLLVP 470
SQE1	SSENEARNEMREACFDYLSLGGMCTSGPVSLLSGLNPRPLTLVCHFFAVAVYGVIRLLIP 476
	:* :*:***:***:***:***: :*: :*: :*: :*: :*: :*: :*: :*: :*: :*

The primary focus was on the SQE1, SQE2 and SQE3, as they have been previously functionally characterised. SQE4, SQE5 and SQE6 all putative Squalene epoxidases closely matched the intensity values for all for plant species; *A. thaliana*, *O. spinosa*, *O. pusilla* and *M. truncatula*. This work has highlighted the need to identify upper limits for intensity values for designing mask files for cross species transcriptional profiling.

Within the transcriptional profile there were clear differences within the expression values of Squalene epoxidases. Within leaf tissue SQE1 was expressed at a higher level than within the root in *O. pusilla* and *O. spinosa*. SQE2 (At2g22830) was highly expressed in *O. pusilla* root and to a lesser extent in leaf. In contrast the expression of SQE2 was lower in *O. spinosa*. The expression levels of SQE3 was the opposite. *O. spinosa* showed high expression levels of SQE3 a 2,3;22,23-dioxidosqualene cyclase in root. SQE3 was also expressed in *O. spinosa* leaf but to a lesser extent.

According to the intensity value from the PM hybridisation of gDNA, a small effect on the expression profile was expected. Encouragingly the expression patterns between *O. pusilla* and *O. spinosa* root and leaf were inverted between the two gene sets suggesting that a gene similar to SQE2 is the primary precursor in *O. pusilla* in both root and leaf. If the hybridisation intensity of SQE3-8 effected the expression value, there may be a similar values between the root and the leaf within *O. spinosa*. This is not the case which suggests a 2,3;22,23-dioxidosqualene synthase has higher



transcriptional value in *O. spinosa* than in *O. pusilla*. The same pattern occurs between *O. pusilla* and *O. spinosa* for SQE5 (At5g24140) and SQE4 (At5g24150), which indicates that these putative squalene synthases may fulfil the same roles as SQE3 and SQE2 respectively. It is worth noting that the PM signal values between all species matched very closely across all probes.

Within the oxidosqualene cyclases there were four main examples where the expression was higher in *O. spinosa* leaf than *O. spinosa* root, *O. pusilla* root and *O. pusilla* leaf. These were At1g66960, At1g78950, At1g78970 and At1g78500, all capable of cyclising  $\beta$ -amyrin. Fundamentally the known 2,3;22,23-dioxidosqualene cyclase enzymes were expressed in high levels with *O. spinosa* root and the  $\beta$ -amyrin synthases were expressed at higher level in *O. spinosa* leaf.

## CHAPTER 7. GENERAL DISCUSSION

### 7.1 ALPHA ONOCERIN AND ITS IMPORTANCE

The cyclisation mechanism for  $\alpha$ -onocerin biosynthesis in plants is currently poorly understood at the molecular and biochemical levels. The stereogenic symmetrical structure, consisting of four cyclic carbon rings coupled by an incomplete central isoprene unit makes for an unusual plant secondary metabolite product. Little is known about the function of  $\alpha$ -onocerin or the role it plays in plant secondary metabolomic pathways in nature. Occurrence across such a diverse range of taxa and association with plants adapted to extreme environmental habitats makes  $\alpha$ -onocerin an intriguing secondary metabolite to study.

Aside from the original elucidation studies performed by Hlasiwetz (1855) and subsequent studies on the derivatives (von Hemmelmayr, 1907; Dieterle, *et al.*, 1934; Zimmermann, 1938 and 1940), studies on the occurrence and cyclisation mechanisms have been few and far between. More recently the structure of  $\alpha$ -onocerin was confirmed using x-ray crystallography by Stork *et al.*, (1963), van Tamelen (1968); Mi *et al.*, (2002), Pauli (1999), Fröhlich and Pauli (2000). Attempts to investigate the occurrence and synthesis of  $\alpha$ -onocerin were carried out by Mike Rowan and Peter Dean (1972a), by studying the occurrence of  $\alpha$ -onocerin in Restharrow. Restharrow occupy a wide range of environments. *O. repens maritima*, *O. reclinata* *O. spinosa* and *O. repens* have been reported to colonise coastal areas. *O. repens maritima* has developed the remarkable ability to withstand highly saline shingle beaches, in areas where the presence of marine flora is very limited. Aside from revealing the occurrence of  $\alpha$ -onocerin in a wide range of Restharrow species that occupy different ecological habitats, the group discovered that  $\alpha$ -onocerin was a product of a squalene enzyme substrate complex.  $\alpha$ -Onocerin is thought to be synthesised by a process infrequently found in nature; cyclisation from both terminals of a 2,3;22,23-dioxidosqualene, (Rowan and Dean, 1972a).

To date there is nothing in the literature to confirm or discount the proposed mechanism by Peter Dean and Mike Rowan (1972a).

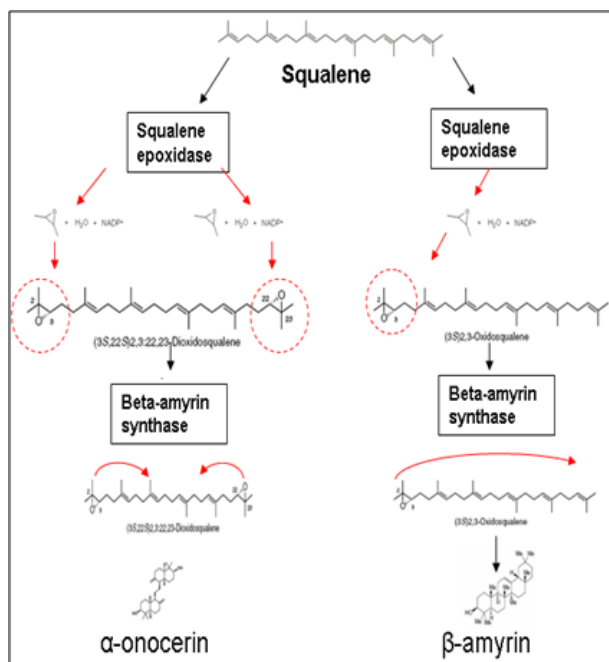
The importance of addressing the biological questions surrounding the biosynthesis of  $\alpha$ -onocerin should not be underestimated.  $\alpha$ -onocerin is an example of a triterpene that has been reported by several groups as having medicinal (Orhan *et al.*, 2005; Zhang *et al.*, 2002; yolmaz *et al.*, 2006), antifungal and antibacterial properties (Mahasneh and El-Oqlah 1999). Research into the medicinal properties of  $\alpha$ -onocerin, as discussed in Chapter 1, has occurred on a few occasions (Mahasneh and El-Oqlah 1999., Orhan *et al.*, 2005, Yolmaz *et al.*, 2006; Rollenger *et al.*, 2003., 2006, Li *et al.*, 2006; Jaric *et al.*, 2007). Traditionally  $\alpha$ -onocerin as a component of plant extracts has been utilised against arthritic pain, dysmenorrhoea, a diuretic, expectorant and an antiseptic (Mahasneh and El-Oqlah, 1999).

Restharrow is known to accumulate large quantities of secondary metabolites within the root membrane (Rowan and Dean 1972b). It is thought  $\alpha$ -onocerin may help by providing a barrier preventing the uptake of sodium chloride ions across the root membrane (Spikett *et al.* 1993) allowing the colonisation of coastal areas and arid regions. Several studies on Restharrow have revealed significant amounts of  $\alpha$ -onocerin accumulate within the root.  $\alpha$ -onocerin is present across a wide range of unrelated plant taxa and associated with plants occupying highly saline or water saturated habitats.

As proposed by Rowan and Dean (1972a) a bi-cyclisation event of 2,3;22,23-dioxidosqualene via an oxidosqualene cyclase may provide plants with an alternative mechanism for biosynthesising a range of triterpene diol products via an  $\alpha$ -onocerin. This may have important implications concerning the ability of plants to adapt to their surrounding environment by biosynthesising alternative water proofing, anti-feedant, antifungal and antimicrobial triterpene saponins. Given appropriate environmental conditions, the ability for plant to produce  $\alpha$ -onocerin may be more widely distributed than previously thought. A relatively simple modification of

existing pathways, may occur simply as a result of changes in transcription of specific squalene epoxidases to cyclase 2,3;22,23-dioxidosqualene. This may offer an alternative cyclisation mechanism towards triterpene diols.

## 7.2 A POTENTIAL BIOSYNTHETIC MECHANISM TOWARDS $\alpha$ -ONOCERIN CYCLISATION



**Figure 7.1** Potential route towards  $\alpha$ -onocerin biosynthesis via an oxidosqualene cyclase. (left)  $\alpha$ -onocerin as a product, via a 2,3;22,23-dioxidosqualene and (right)  $\beta$ -amyryn as a product of 2,3-oxidosqualene.

There are three proposed potential routes for  $\alpha$ -onocerin biosynthesis. The first mechanism could involve a specific oxidosqualene cyclase enzyme responsible for the cyclisation ( $\alpha$ -onocerin cyclase). A second mechanism is based on the proposal from Rowan and Dean (1972a) that a single oxidosqualene cyclase processes a 2,3;22,23-dioxidosqualene. The third mechanism proposed an alternative sub-pathway, via squalene as an initial precursor. During the course of the project it was ascertained that there was sufficient evidence to discount hypothesis 3.

### **7.3 GENEVESTIGATOR CAN BE USED AS A TOOL TO IDENTIFY TRANSCRIPTIONAL CHANGES ACROSS BIOCHEMICAL PATHWAYS.**

A summary of the genes involved within the triterpene pathway, was deduced by investigating the transcriptional changes in the model crop *A. thaliana*. Arabidopsis Gene Initiative (AGI) codes and Affymetrix GeneChip®, ATH-1 probe set identification numbers were identified (table 3.1). Information about the Genes involved in triterpene biosynthesis was ordered according to their involvement in each stage within the pathway. Individual oligonucleotide sequences were searched against the Genbank database using the online bulk search sequences from the Affymetrix NetAffx ([www.affymetrix.com](http://www.affymetrix.com)). These tools were used as an initial assessment of cross hybridisation between individual oligonucleotide probes. Heat maps were generated using transcriptional expression signal profiles of *A. thaliana* across the triterpene pathway within GENEVESTIGATOR. SQS1 (At4g34640) had higher expression than SQS 2 (At4g34650), suggesting it may be the primary gene involved in encoding a functional squalene synthase.

At the individual stage level, within *A. thaliana*, At5g24150 (SQE5) and At4g37760 (SQE3), were consistently expressed at a higher level than the other homologues throughout development. SQE1 (At1g58440) and SQE3 (At4g37760), were both expressed at a higher level than the other homologues at early development stages. SQE1 and SQE3 have been shown to cyclise 2,3-22,23-dioxidosqualene. This may suggest 2,3;22,23-dioxidosqualene is important in triterpene biosynthesis during early plant development. Alternatively SQE1 and SQE3 be simply be cyclising 2,3-oxidosqualene as the primary substrate as discussed in Chapter 3.

#### **7.4 PUBLICALLY AVAILABLE INFORMATION MAY OFFER AN INSIGHT INTO POTENTIAL TARGETS FOR $\alpha$ -ONOCERIN BIOSYNTHESIS INVESTIGATION**

The main focus of Chapter 4 was to use genome and metabolome profiling tools available online. The information gathered from these results were utilised as a starting point for designing future experiments. Initially there was very little data available concerning the triterpene pathway in *Restharrow*. No data was available other than that squalene was a precursor of  $\alpha$ -onocerin. As a result, a metabolomic map was developed from the KEGG database. Combined searches outlined key information combining both metabolomic and genomic data sets summarised in a pathway context. No genomic sequence information from *Restharrow* is currently available. Aracyc was utilised to summarise genomic information provided by the Plant Metabolic Network. This in turn allowed selection of biochemical markers to investigate triterpene biosynthesis using metabolomics.

*A. thaliana* was chosen as a model, due to the availability of genetic information to investigate and map the genes involved in the triterpene pathway. In a pathway context; in *A. thaliana* there are two copies of squalene synthase; At4g34640 and At4g34650, Six squalene epoxidases At1g58440, At2g22830, At4g37760, At5g24140, At5g24150, and At5g24160. Four oxidosqualene cyclases were present in *A. thaliana* situated as a gene cluster on chromosome 1; At1g66960 lupeol synthase, At1g78950  $\beta$ -amyrin synthase, At1g78960 lupeol synthase and At1g78970 lupeol synthase. A total number of six other triterpene synthases were present; At1g78500, At4g15340, At4g15370, At5g36150, At5g42600 and At5g48010. A high level of conservation exists between gene homologues, suggesting gene duplications may have occurred. This interpretation is supported by a large number of clustered homologues being present in chromosomal locations within the Arabidopsis genome. Squalene synthase is the initial precursor to all terpenes, within *Medicago truncatula* there is a single copy within the genome. It is expected that in *Restharrow* there may be a single copy of squalene synthase as they are both legumes. This result

is supported by the situation in *Panax gensing*, where a single copy of squalene synthase is present.

Within the Aracyc database a limited amount of information was available for squalene epoxidases. Four homologues of squalene epoxidase were described in *Solarium lycopersicum*, and six in *A. thaliana*. Of which three squalene epoxidases have been functionally characterised using over expression studies in yeast; At1g58440, and At4g37760. Importantly At1g58440 and At4g37760 have been shown to cyclise a 2,3;22,23-dioxidosqualene either as a primary product or as a bi-product of 2,3-oxidosqualene biosynthesis. In contrast 2,3;22,23-dioxidosqualene was not detected in yeast cells where SQE2 (At2g822830) was over-expressed. A single SQE amplified from Restharrow using degenerate oligonucleotide pairs designed from the consensus amino acid sequences of squalene epoxidases. From the results of amino acid BLAST searches and alignments, a higher level of conservation of the squalene epoxidase isolated from Restharrow was seen between SQE3 (At4g37760). The other five *A. thaliana* SQE showed reduced homology with the *Ononis* version.

## **7.5 AMINO ACID SEQUENCE HOMOLOGY MAY SUGGEST FUNCTION OF PUTATIVE OXIDOSQUALENE CYCLASE ISOLATED FROM RESTHARROW**

Within *Pisum sativum*, closely evolutionarily related to Restharrow, there were single copies of  $\beta$ -amyrin synthase (BAA97558; PSY) and a single multi-functional triterpene synthase gene; (BAA97559; PSM). A partial putative oxidosqualene cyclase amino acid sequence, isolated from Restharrow by Sean Mayes (unpublished data), was extended using rapid amplification of cDNA ends to include three Arginine, a single Valine, Proline, Leucine and Proline (RRRVPLP-3') residues. The putative oxidosqualene cyclase amino acid sequence was subsequently aligned (ClustalW2, EBI), with the PSY and PSM *P. sativum*.

The amino acid sequence within the active site of all three proteins was highly conserved; Aspartic acid, Cystine, Theonine, Alanine and Glutamic acid (DCTAE). A region of the amino acid sequence thought to be involved in triterpene specificity showed differentiation between the *P. sativum* PSM and PSY. The Tryptophan residue (W) in the sequence; Methionine, Tryptophan, Cystine, Tyrosine, Cystine and Arginine (MWCYCR) was identical between the known  $\beta$ -amyrin synthase in *P. sativum* (PSY) and the Restharrow putative oxidosqualene cyclase. The amino acid alignments showed greater sequence homology and structural similarity between the Restharrow putative-oxidosqualene cyclase and the *P. sativum*  $\beta$ -amyrin synthase (BAA7558; PSY), than the *P. sativum* multifunctional oxidosqualene cyclase (PSM).

The multifunctional oxidosqualene cyclase (PSM), sequence from *P. sativum* had a Leucine in the equivalent position MLCYCR. Although Tryptophan and Leucines are both non-polar, Leucine has a positive charge whereas, Tryptophan has a negative charge. This single amino acid difference may affect the product specificity (Husselstein-Muller *et al.*, 2001; Lee *et al.*, 2004). Based on the *in silico* studies, amino acid alignments, phylogenetic analysis and 3D hypothetical modelling, the Restharrow putative triterpene synthase is likely to be a  $\beta$ -amyrin synthase, rather than a specific  $\alpha$ -onocerin cyclase. Over-expression studies would be required for full functional characterisation of the Restharrow putative  $\beta$ -amyrin synthase.

## **7.6 BIOCHEMICAL EVIDENCE TOWARDS $\alpha$ -ONOCERIN VIA A 2,3;22,23-DIOXIDOSQUALENE, CYCLISATION ROUTE**

Based on the biochemical evidence presented in this thesis, hypothesis 2 is the most likely mechanism to occur in Restharrow for the biosynthesis of  $\alpha$ -onocerin. Both 2,3;22,23-dioxidosqualene and 2,3-oxidosqualene were present in *O. spinosa* where  $\alpha$ -onocerin was detected as an accumulated product in plant tissue. In contrast, 2,3;22,23-dioxidosqualene and  $\alpha$ -onocerin could not be detected in enzyme whole cell preparations, leaf



and root and developmental time points in both *O. pusilla* and *O. rotundifolia*. Within three species *O. spinosa*, *O. repens maritima* and *O. repens* 2,3;22,23-dioxidosqualene was present wherever  $\alpha$ -onocerin was detected. In *O. spinosa*, *O. repens* and *O. repens maritima* there were no detectable levels of 2,3;22,23-dioxidosqualene or  $\alpha$ -onocerin within leaf extracts. This not only suggests that the biosynthetic mechanism for  $\alpha$ -onocerin is present within the roots, but that the 2,3;22,23-dioxidosqualene is biosynthesised primarily within the root. Cyclisation may alternatively occur within the leaf, however is less stable or in undetectable quantities and subsequently transported to the root. By detecting the presence of 2,3-oxidosqualene, 2,3;22,23-dioxidosqualene,  $\alpha$ -onocerin and  $\beta$ -amyrin, the biochemical evidence supports hypothesis 2. The biochemical evidence suggests the control of  $\alpha$ -onocerin cyclisation would occur at the squalene epoxidase level therefore being substrate specific. The re-entry of 2,3-oxidosqualene into the active site of the squalene epoxidase may be the rate limiting step and offer a feedback mechanism for triterpene biosynthesis in Restharrow. In contrast there was little evidence to support hypothesis one as 2,3;22,23-dioxidosqualene was present and a direct trade-off between  $\alpha$ -onocerin and  $\beta$ -amyrin accumulation was not seen. The fact that a limited quantity of 2,3-oxidosqualene was available in plant extracts when  $\alpha$ -onocerin was in abundance, suggests the 2,3;22,23-dioxidosqualene was the primary substrate used. By observing the data in figure 5.39 and 5.40, it is evident that 2,3;22,23-dioxidosqualene was present at greater levels than 2,3-oxidosqualene. Subsequently the level of  $\alpha$ -onocerin was significantly greater than  $\beta$ -amyrin. This provides evidence that the availability of substrate may affect the accumulation of  $\alpha$ -onocerin or  $\beta$ -amyrin, which is controlled via a single oxidosqualene cyclase likely to be similar to the putative  $\beta$ -amyrin synthase as outlined in chapter 4.

## **7.7 AFFYMETRIX ATH-1 CAN BE USED FOR CROSS SPECIES TRANSCRIPTOMICS OF RESTHARROW *O. PUSILLA* AND *O. SPINOSA*.**

Transcriptional profiling using cross species hybridisation on an Affymetrix GeneChip (X-Species) is an accepted method of transcriptional profiling of organisms where dedicated GeneChip technology is not available. Triterpene synthases for *O. spinosa* and *O. pusilla* root and leaf RNA were transcriptionally profiled on the ATH-1 GeneChip. *O. spinosa* gDNA hybridisation signal value was normalised using RMA analysis using cross species hybridisation discovery software, PIGEON (Unpublished, Mayes and Lai, (2010)). The threshold of probe pairs against probes sets lost at thresholds 20 to 1000 intensity were plotted. For *O. spinosa* gDNA the threshold 100 was used to generate a masking file, also using PIGEON. At this threshold the number of probe sets retained was 20544 and probe pairs was 55543. Thus 90.75% of probes were retained and 22.61% of probe pairs were retained. This was considered an acceptable degree of probe pair loss in comparison to the number of probes retained. In *O. pusilla* the number of probe sets retained was 20567 (90.41%) and probe pairs retained was 56500 (22.59%) at signal threshold of 100.

## **7.8 PERFECT MATCH AND MISMATCH DATA OF INDIVIDUAL OLIGONUCLEOTIDES CAN BE USED TO IDENTIFY CROSS HYBRIDISATION**

The gDNA hybridisation threshold data was examined. Generally gDNA from *A. thaliana* should theoretically have the highest signal values using the Perfect Match oligonucleotides. This was not always the case across the twenty-one genes involved in triterpene biosynthesis. Across the probe set signal values from *M. truncatula*, *O. pusilla* and *O. spinosa*, the same pattern as the *A. thaliana* gDNA was seen. Differences in signal may be explained by the differences in hybridisation efficiencies of the individual experiments and differences between nucleotide sequences.

The gDNA Perfect Match signal values were used to identify potential sequence variation between *O. pusilla* and *O. spinosa*. Within the squalene epoxidase SQE1 oligonucleotide 6 (SQE1-6) known to cyclise 2,3;22,23-dioxidosqualene) showed differences between *M. truncatula* and *O. pusilla*. The highest intensity was 272 in *O. spinosa* and 157 in *A. thaliana*, whereas in *M. truncatula* and *O. pusilla* the intensity was 13 and 30 respectively. The intensity values for the other ten oligonucleotides for SQE1 matched closely. While it is not possible to know the nature of the polymorphism in the absence of sequence information for all of the gene equivalents, it might suggest polymorphism that could have a functional effect. This may lead to functional differences between *O. pusilla* and *O. spinosa*. For squalene epoxidase SQE2-3, *O. pusilla* had a signal value 784 which may represent a difference in sequence or a cross hybridisation onto an alternative oligonucleotide. As a result SQE2-3 was disregarded.

The SQE3-5 in *M. truncatula* the intensity was 93, whereas *A. thaliana* was 8; this may equate to a sequence difference. SQE3-8 gave a signal value of 177 in *O. spinosa* in contrast *O. pusilla* had an intensity of 40 and *A. thaliana* had a signal intensity of 9. Oligonucleotide SQE3-8 may indicate that sequence variation between *O. pusilla* and *O. spinosa* is present.

**Table 7.1** The oligonucleotide sequences where high intensity values occurred were; SQE1-6, SQE2-3, SQE3-8. The gene and the nucleotide sequence are listed.

Gene	Nucleotide sequence
SQE1-6	GAACCCTCGACCATTAACACTTGTC
SQE2-3	AGGCCAATTCGCAACCTTAATGACA
SQE3-8	CATCAGGACCTGTGGCTTTGCTATC

These sequences were translated into amino acid sequences and aligned using ClustalW to the Arabidopsis thaliana sequence (below) and other available sequences ([www.EBI.ac.uk](http://www.EBI.ac.uk))



## 7.9 TRANSCRIPTIONAL EVIDENCE TOWARDS $\alpha$ -ONOCERIN CYCLISATION ROUTE VIA A 2,3;22,23-DIOXIDOSQUALENE CONTROLLED BY SQUALENE EPOXIDASE

Across transcriptional profiles there were clear differences in the expression values of squalene epoxidases. In leaf tissue SQE1 was expressed at a higher level than within the root in *O. pusilla* and *O. spinosa*. SQE2 (At2g22830) was highly expressed in *O. pusilla* root and to a lesser extent in *O. pusilla* leaf. In contrast the expression of SQE2 was lower in *O. spinosa* than in *O. pusilla* root. The high level of expression for SQE3 (a 2,3;22,23-dioxidosqualene cyclase), in *O. spinosa* root, was also replicated in *O. spinosa* leaf.

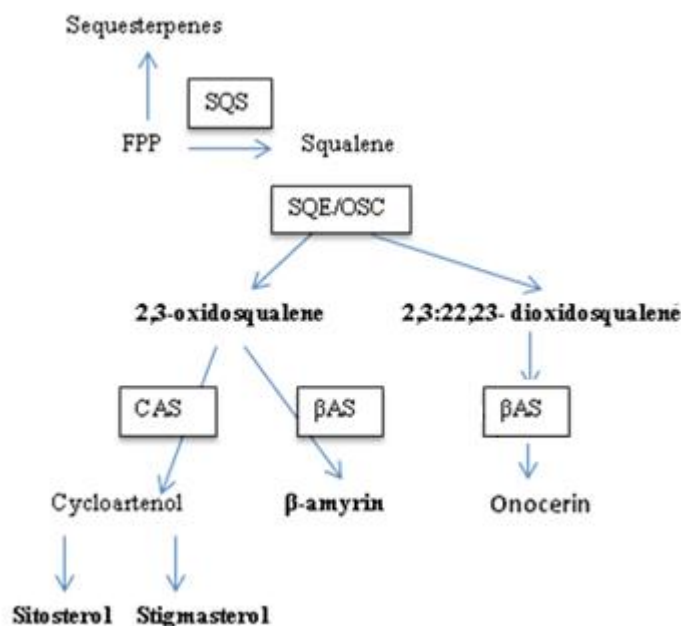
Within the oxidosqualene cyclases there were four cases where the expression appeared higher in *O. spinosa* leaf than *O. spinosa* root, *O. pusilla* root and *O. pusilla* leaf. These were At1g66960, At1g78950, At1g78970 and At1g78500, all capable of cyclising  $\beta$ -amyrin. The known 2,3;22,23-dioxidosqualene cyclase enzymes were expressed at high levels in *O. spinosa* root and the  $\beta$ -amyrin synthases were expressed at higher levels in *O. spinosa* leaf.

The transcriptional profiling confirmed that although there may be some sequence differences between *A. thaliana* and *Ononis species*, 2,3;22,23-dioxidosqualene synthase (producing squalene bis-epoxidase) is primarily expressed within the roots of *O. spinosa*. It is clear that there are some fundamental differences between the transcriptional profiles of *O. pusilla* and *O. spinosa*. SQE2 was more highly expressed in *O. pusilla*, suggesting the production and use of 2,3-monoepoxide. Summarised in a figure 7.2, a diagrammatic representation is shown, which outlines a modified  $\alpha$ -onocerin cyclisation mechanism in Restharrow, based on the evidence presented in the thesis.

## 7.10 SUMMARY: A PROPOSED MECHANISM FOR $\alpha$ -ONOCERIN BIOSYNTHESIS

As deduced from a number of comparative studies; triterpene pathway information in *A. thaliana*, genomic information and metabolomic evidence and cross species transcriptional profiling,  $\alpha$ -onocerin biosynthesis is likely be biosynthesised by the proposed route outlined in figure 7.2. A summary of the primary observations from work presented in this thesis below:

1. A squalene epoxidase cyclising 2,3;22,23-dioxidosqualene, from a 2,3-oxidosqualene re-entering the active site, prior to a cyclisation event.
2. A cyclisation event by an oxidosqualene cyclase likely to be a  $\beta$ -amyrin synthase on 2,3;22,23-dioxidosqualene (Hypothesis 2).
3. The cyclisation event likely to occur within the roots of restharrow, which indicates  $\alpha$ -onocerin may play a functional role within the roots rather than the aerial tissue.
4. Evidence to suggest 2,3;22,23-dioxidosqualene and  $\alpha$ -onocerin doesn't occur regularly at detectable levels in *O. pusilla*.
5. A specific squalene bis-epoxidase similar to SQE3, is likely to be the primary enzyme responsible for cyclising the 2,3;22,23-dioxidosqualene as the primary precursor.
6. Low transcription level or absence of a squalene bis-epoxidase in *O. pusilla* is likely to affect the plants ability to cyclise  $\alpha$ -onocerin.



**Figure 7.2** Triterpene biosynthetic pathway within the root of Restharrow *O. spinosa*, *O. repens* and *O. repens*, subsp-*maritima*. Farnesyl diphosphate (FPP), squalene synthase (SQS), squalene epoxidase (SQE/OSC), (cycloartenol synthase (CAS) and  $\beta$ -amyrin synthase ( $\beta$ AS).

## 7.11 FUTURE DIRECTIONS

In Restharrow, there are likely to be two squalene epoxidase, one  $\beta$ -amyrin synthase and one multi-functional triterpene synthase, as in *Pisum*. This work has identified that the key genes that are most likely to affect  $\alpha$ -onocerin biosynthesis in Restharrow are squalene epoxidases. The work has helped develop microarray and metabolomic methods for future functional studies.

Genome sequencing has become more readily available and cost effective. Proposed directions for future projects would focus on sequencing the genomes of *O. spinosa* and *O. pusilla*. Other Restharrow species may also be considered such as *O. repens* subsp. *maritima*, *O. repens* and *O. rotundifolia*. From this, a series of BLAST searches, bioinformatic, alignments and phylogenetics could be carried out for comparative genomic studies.

Sequence alignments at the amino acid level may be used for identifying squalene epoxidase and oxidosqualene cyclase homologues, between species likely to be involved in the pathway. These targets namely squalene epoxidase and oxidosqualene cyclase may be selected for functional characterisation studies. These genes from each species could be synthesised and over-expressed *Saccharomyces* or *Arabidopsis thaliana*. Methods developed within the thesis could be used evaluated functional characterisation studies.

One future experiment could utilise these methods by looking at the effect changing environmental conditions have on the transcriptional profile and metabolomic accumulation. Such studies may include saline and water saturation experiments.

The project has also generated a plethora of micro-array data, which may be utilised in future studies. One suggestion includes the analysis of microarray data, between *O. pusilla* and *O. spinosa*, comparing changes in gene expression values across genes involved within the phytosterol pathway.

Further research into the occurrence of  $\alpha$ -onocerin across a diverse range of taxa including club mosses and ferns and plants known to biosynthesise triterpene diols, may compliment the research outlined above using methods developed within this thesis.



## REFERENCES

- ABE, I., ABE, T., LOU, W., MASUOKA, T., & NOGUCHI, H. (2006) Site-directed mutagenesis of conserved aromatic residues in rat squalene epoxide *Biochemical and Biophysical Research Communications*, 352, 259-263.
- ABE, I., & PRESTWICH, G.D. (1999) Squalene epoxidase and oxidosqualene lanosterol cyclase; key enzymes in cholesterol biosynthesis. In: *Comprehensive Natural Products Chemistry*, pp. 267-298. Elsevier, New York.
- AGETA, H., IWATA, K., & OOTAKE, Y. (1962) Isolation of alpha-onocerin from *Lycopodium clavatum* Linn. *Chemical and Pharmaceutical Bulletin, (Tokyo)*, 10, 637.
- AGETA, H., MASUDA, K., & INOUE, M. (1982) Fern constituent - colysanoxide, and onoceroid having novel carbon skeleton, isolated from colysis species. *Tetrahedron Letters*, 23 (42), 4349-4352.
- ANDERSON, L., & SEILHAMER, J. (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis*, 18, 533-7.
- ARGUESO, C.T., FERREIRA, F.J., & KIEBER, J. (2009), Environmental perception avenues: the interaction of cytokinin and environmental response pathways. *Plant, Cell & Environment*, 32, 1147-1160.
- ARONSON, J., KIGEL, J., SHMIDA, A., (1992) Reproductive allocation strategies in desert and Mediterranean populations of annual plants grown with and without water stress. *Oecologia*, 93:3, 336-342.
- AUBERG, S., LECHAMY, A., & BOHLMANN, J. (2002) Genomic analysis of the triterpenoid synthase (AtTPS) gene family of *Arabidopsis thaliana*. *Molecular Genetics and Genomics*, 276:6, 730-745.
- AUGENLICHT, L.H., WAHRMAN, M.Z., HALSEY, H., ANDERSON, L., TAYLOR, J., & LIPKIN, M. (1987) Expression of cloned sequences in biopsies of human colonic tissue and in colonic carcinoma cells induced to differentiate *in vitro*. *Cancer Research*, 47, 6017.
- BALANDRIN, M.F., KLOCKE, J.A., WURTELE, E.S. & BOLLINGER, W.H. (1985) Natural plant chemicals: Sources of industrial and medicinal materials. *Science*, 228, 1154-60.
- BANDARANAYAKE, W.M. (2002) Bioactive compounds and chemical constituents of mangrove plants. *Wetlands Ecology and Management*, 10, 421-452.
- BARRETT-LENNARD, E.G. (2003) The interaction between water logging and salinity in higher plants: causes, consequences and implications *Plant and Soil*, 253, 35-54.

- BARTON, D.H.R., JARMAN, T.R., WATSON, K.C., WIDDOSON, D.A., BOAR, R.B., & DAMPS, K. (1975) *Journal of the Chemical Society*, 1, 1134-1138.
- BARTON, D.H.R., & OVERTON, K.H. (1955) Constitution and stereochemistry of onocerin. *Chemistry and Industry*, 23, 654-655.
- BATANOUNY, K.H. (2001) Plants in the deserts of the Middle East. *Springer-Verlag*. Berlin Heidelberg New York.
- BENNETT, R.N., & WALLSGROVE, R.M. (1994) Secondary metabolites in plant defence mechanisms. *New phytologist*. 127;617-633
- BHUTANI, K.K., KAPOOR, R., & ATAL, C.K. (1984) Two unsymmetric tetracyclic triterpenoids from *Cissus quadrangularis*. *Phytochemistry*, 23, 407-410.
- BIASTED, D.J. (1971) Sterol and triterpene synthesis in the developing and germinating pea seed. *Journal of Biochemistry*, 124, 375-383.
- BISSET, M. (2001): a handbook for practice on a scientific basis; with reference to German Commission E monographs. In *Herbal drugs and phytopharmaceuticals 2<sup>nd</sup> edition*. Medpharm Scientific Publishers.
- BLOCH, K. (1992) Sterol molecule: structure, biosynthesis and function. *Steroids*, 57, 378-383.
- BOITEAU, P., PASICH, D., & RATSINMAMANGA, A.R. (1962) Les triterpenoids. *Gauthier-villars, Paris* 1964.
- BOUVIER, P., BERGER, Y., ROHMER, M., OURISSON, G. (1980) Non-specific biosynthesis of gammacerane derivatives by a cell-free system from protozoan *Tetrahymena pyriformis*. Conformations of squalene, (3s)-squalene epoxide and (3R)-squalene epoxide during cyclisation. *European Journal of Biochemistry*, 112:3, 549-56.
- BRAVO, L. (1998) Polyphenols: Chemistry, Dietary Sources, Metabolism, and Nutritional Significance. *Nutrition Reviews*, 56; 11, 317-333.
- BRENNER, S., JOHNSON, M., BRIDGHAM, J., GOLDA, G., LLOYD, D.L., LUO, S., MCCURDY, S., FOY, M., EWAN, M., ROTH, GEORGE, D., ELETR, S., ALBRECHT, G., VERMASS, E., WILLIAMS, S.R., MOON, K., BURCHAM, M.P., DUBRIDGE, R.B., KIRCHNER, J., FEARON, K., MAO, J.I., & CORCORAN, K. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *National Biotechnology*, 18(6), 630-634.
- BROADLEY, M.R., WHITE, P.J., HAMMOND, J.P., GRAHAM, N.S., BOWEN, H.C., EMMERSON, Z.F., FRAY, R.G., IANNETTA, P.P.M., MCNICOL, J.W., MAY, S.T., (2008) Evidence of neutral transcriptome evolution in plants. *New Phytologist*, 180, 587-593.

- BULLOCK, J.A. (1992) Host plants of British beetles: A list of recorded associations. *Amateur Entomologists Society*, 11a, 24.
- BUSTIN, S.A. (2002) Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *Journal of Molecular Endocrinology*, 29, 23-39.
- CAI, X., PAN, D.J., XU, G.Y., SETO, H., & FURIHATA, K. (1992) Novel pentacyclic triterpene esters of serratene type from *Lycopodium obscurum* L. *Acta Chimica Sinica*, 50, 60-66.
- COREY, E.J., CHENG, H., BAKER, C.H., MATSUDA, S.P.T., LI, D., & SONG, X. (1997) Methodology for the preparation of pure recombinant *S. cerevisiae* lanosterol synthase using a baculovirus expression system. Evidence that oxirane cleavage and A-ring formation are concerted in the biosynthesis of lanosterol from 2,3-oxidosqualene. *Journal of the American Chemical Society*, 119, 1277-1288.
- COREY, E.J., & GROSS, S.K.J. (1992) A lanosterol synthase prefers DOS over OS *American Chemical Society*, 89, 4561-4562.
- CRAIGON, D.J., JAMES, N., OKYERE, J., HIGGINS, J., JOTHAM, J., & MAY, S. (2004) NASCArrays: a repository for microarray data generated by NASC's transcriptomics service. *Nucleic Acids*, 32, 575-577.
- CREELMAN, R.A., & MULLET, J.E. (1995) Jasmonic acid distribution and action in plants: regulation during development and response to biotic and abiotic stress. *Proceedings of the National Academy of Sciences*, 92, 4114-4119.
- CRESPO, M.B. (1991) A new species of the genus *Ononis* L. (Fabaceae) *Botanical Journal of the Linnean Society*, 111, 37-46.
- CRICK, F. (1970) Central Dogma of Molecular Biology. *Nature*, 227, 561-563.
- DARUHAZI, A.E., SZARKA, S., HETHELYI, E., SIMANDI, B., GYURJAN, I., AND LEMBERKOVICS, E. (2008) GC-MS identification and GC-FID Quantitation of Terpenoids in *Ononidis spinosae radix*. *Chromatographia*, 68, 71-76.
- DAVEY, M.W., KEULEMANS, J., GRAHAM, N., MAY, S.T., VANHOLME, B., & SWENNEN, R. (2009) Assessing the use of heterologous oligonucleotide microarrays for transcriptomics in a non-model species: application to the study of drought stress in *Musa*. *Acta Horticulturae*, 839, 607-613.
- DAVISS, B. (2005). "Growing pains for metabolomics". *The Scientist*, 19, 25-28.

- DELLA GRECA, M., FIORENTINO, A., MONACO, P., & PREVITERA, L. (1994) Cycloartane triterpenes from *Juncus effusus*. *Phytochemistry*, 35, 1017–1022.
- DE PASCUAL TERESA, J., URONES, J.G., MARCOS, I.S., BASABE, P., SEXMERO CUADRADO, M.J., FERNANDEZ & MORO, R. (1987) Triterpenes from *Euphorbia broteri*. *Phytochemistry*, 26, 1767–1776.
- DEMEL, R.A., DE KRUFF, B. (1976) The functions of sterols in membranes. *Biochemical Physics*, 1976, 457.
- DE RISI, J., PENLAND, L., BROWN, P.O., BITTNER, M.L., MELTZER, P.S., RAY, M., CHEN, Y., SU, Y.A., & TRENT, J.M. (1996) Use of a cDNA microarray to analyze gene expression patterns in human cancer. *Nature*, 14, 457–460.
- DE RISI, J.L., IYER, V.R., & BROWN, P.O. (1997) Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale. *Science*, 278, (5338), 680–686
- DE VOSS, J.J., SIBBESEN, O., ZHANG, Z.P., & ORTIZ DE MONTELLANO, P.R. (1997) Substrate Docking Algorithms and Prediction of the Substrate Specificity of Cytochrome P450cam and its L244A Mutant. *Journal of the American Chemical Society*, 119, 5489–5498.
- DIETERLE L.H., SALOMOM, A., & GARTNER, R. (1934) *Archive der Pharmazie*, 272, 142.
- DIXON, R.A., GANG, D.R., CHARLTON, A.J., & FIEHN, O. (2006) Applications of metabolomics in agriculture. *Journal of Agricultural Food*, 54 (24), 8984–94.
- EDGAR, R., DOMRACHEV, M., & LASH, A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*, 30, 207–210.
- FARAG, M.A., HUHMANN, D.V., DIXON, R.A. & SUMNER, L.W. (2008) Metabolomics reveals novel pathways and differential mechanistic and elicitor-specific responses in phenylpropanoid and isoflavonoid biosynthesis in *Medicago truncatula* cell cultures. *Plant Physiology*, 146: 2, 387–402.
- FIELD, R.B., & HOLMLUND, C.E. (1977) Isolation of 2,3;22,23-dioxidosqualene and 24,25-oxidolanosterol from yeast. *Archives of Biochemistry and Biophysics*, 30, 465–471.
- FRÖHLICH, R., & PAULI, G.F. (2000)  $\alpha$ -Onocerin chloroform hemisolvate. *Acta Crystallographica*, C56, 1476–1477.
- FUJITA, S., OHINISHI, T., WATANBE, B., YOKOTA, T., TAKATSUTO, S., FUJIOKA, S., YOSHIDA, S., SAKATA, K., & MIZUTANI, M. (2006) *Arabidopsis* CYP90B1 catalyses the early C-22 hydroxylation of C<sub>27</sub>, C<sub>28</sub> and C<sub>29</sub> sterols. *Plant Journal*, 45, 765–774.

FULTON, D.C., KROON, P.A., THRELFALL, D.R. (1994) Enzymological aspects of the redirection of triterpenoid biosynthesis in elicitor-treated cultures of *Tabernaemontana divaricata*. *Phytochemistry* 35, 1183-1183.

GANG, D.R., WANG, J., DUDAREVA, N., HEE, N.K., SIMON, J.E., LEWINSOHN, E., & PICHERSKY, E. (2001). An investigation of the storage and biosynthesis of phenylpropenes in sweet basil. *Plant Physiology*, 125, 539–555.

GARFINKEL, D. (1958) Studies on pig liver microsomes. I: enzymic and pigment composition of different microsomal fractions. *Archives of Biochemistry and Biophysics*, 77, 493-509.

GIANNASI, D.E., & NIKLAS, K.L. (1985) The paleobiochemistry of fossil angiosperm floras Part 1. Chemoecystematic aspects. In: Smiley, C.J. (Edn.), Late Cenozoic history of the Pacific North West. *Proceedings of the American Association for the Advancement of Science Meeting. Pacific Division San Francisco*, 173-183.

GOUW, J.W., KRIJGSVELD, J., & HECK, A.J.R. (2009) Quantitative proteomics by metabolic labeling of model organisms. *The American Society for Biochemistry and Molecular Biology*, 9, 11-24.

GRUNWALD, C. (1975) *Phytochemistry*, 14, 75.

GUBLER, U., & HOFFMAN, B.J. (1983) A simple and very efficient method for generating cDNA libraries. *Gene*, 25, 263-69.

GUHLING, O., HOBL, B., YEATS, T., & JETTER, R. (2006) Cloning and characterisation of lupeol synthase involved in the synthesis of epicuticular wax crystals on stem and hypocotyls surfaces of *Ricinus communis*. *Archives of Biochemistry and Biophysics*, 448, 60-72.

GUENGERICH, F.P. (1991) Reactions and significance of cytochrome P450 enzymes. *Journal of Biological Chemistry*, 30 (43), 10513-22.

GUENGERICH, F.P., & JOHNSON, W.W. (1997) Kinetics of ferric cytochrome P450 reduction by NADPH-cytochrome P450 reductase: rapid reduction in the absence of substrate and variations among cytochrome P450 systems. *Biochemistry*, 36 (48), 14741-50.

GUY, C., KOPKA, J., & MORITZ, T. (2008) Plant metabolomics coming of age. *Physiologia Plantarum*, 132, 113-6.

HAAS, B.J., WORTMAN, J.R., RONNING, C.M., HANNICK, L.I., SMITH JR., R.K., MAITI, R., CHAN, A.P., YU, C., FARZAD, M., & WU, D. (2005) Complete reannotation of the Arabidopsis genome: Methods, tools, protocols and the final release. *BioMed Central Biology*, 3, 7.

HABAGUCHI, K., WATANABE, M., NAKADIARA, Y., NAKANISHI, K., KIANG, A.K., & LIM, F.Y. (1968) The full structures of linsic acid and

it's minor congener, an unsymmetric onoceradienedione. *Tetrahedron Letters*, 9, 3731-3734.

HALL, A.E., FIEBIG, A., & PREUSS, D. (2002) Beyond the Arabidopsis genome: Opportunities for comparative genomics. *Plant Physiology*, 129, 1439-1447.

HAMMOND, J.P., BOWEN, H.C., WHITE, P.J., MILLS, V., PYKE, K.A., BAKER, A. J.M., WHITING, S N., MAY, S.T., & BROADLEY, M.R. (2005) A comparison of the *Thlaspi caerulescens* and *Thlaspi arvense* shoot transcriptomes. *New Phytologist*, 170, 239–260.

HAMMOND, J.P., BROADLEY, M.R., CRAIGON, D.J., HIGGINS, J., EMMERSON, Z.F., TOWNSEND, H.J., WHITE, P.J., & MAY, S.T. (2005) Using genomic DNA-based probe-selection to improve the sensitivity of high-density oligonucleotide arrays when applied to heterologous species. *Plant Methods*, 1, 10.

HARALAMPIDIS, K., BRYAN, G., QI, X., PAPADOPOULOU, K., BAKHT, S., MELTON, R., & OSBOURN, A. (2001) A new class of oxidosqualene cyclases directs synthesis of antimicrobial phytoprotectants in monocots. *Proceeding of the National Academy of Sciences*, 98 (23), 13431-13436.

HARALAMPIDIS, K., TROJANOWSKA, M., & OSBOURN, A.E. (2002) Biosynthesis of triterpenoid saponins in plants. *Advances in Biochemical Engineering /Biotechnology*, 75, 81-49.

HARBORNE, J.B. (1999) The comparative biochemistry of phytoalexin induction in plants. *Biochemical systematics and Ecology*, 27,335-367.

HASEGAWA, P.M. & BRESSAN, R.A. (2000) Plant cellular and molecular responses to high salinity. *Annual Review of Plant Physiology and Plant Molecular Biology*, 51, 463.

HAYASHI, H., HUANG, P., KIRAKOSYAN, A., INOUE, K., HIRAOKA, N., IKESHIRO, Y., KUSHIRO, T., SHIBUYA, M., & EBIZUKA, Y. (2001) Cloning and characterisation of cDNA encoding  $\beta$ -amyrin synthase involved in glycyrrhizin and soyasaponin biosynthesis in liquorice. *Biological Pharmacology Bulletin*, 34, (8) 912-916.

HERRERA, J.B.R., BARTEL., WILSON, W.K. & MATSUDA, S.P.T. (1998) Cloning and the characterisation of the *Arabidopsis thaliana* lupeol synthase gene. *Phytochemistry*, 49, 1905-1911.

HLASIWETZ, H. (1855) *Journal of practical Chemistry*, 65, 419.

HOGEWEG, P., SEARLS, D.B. (2011) The Roots of Bioinformatics in Teroretical Biology. *PLOS Computational Biology*, 7(3), e1002021.

- HOLTVOETH, J., VOGEL, H., WAGNER, B., AOLFF, G.A. (2010) Lipid biomarkers in Holocene and glacial sediments from ancient lake Ohrid (Macedonia, Albania). *Biogeosciences*, 7:11, 3473-3489.
- HOSTETTMANN, K.A., & MARSTON, A. (1991) *Saponins*. Cambridge University Press, Cambridge, UK.
- HUSSELSTEIN-MULLER, T., SCHALLER, H., & BENVENISTE, P. (2001) Molecular cloning and expression in yeast 2,3-oxidosqualene triterpenoid cyclases from *Arabidopsis thaliana*. *Plant Molecular Biology*, 45, 75-92.
- INUBUSHI, Y., HARAMAYA, T., HIBINO, T., & AKATSU, M. (1971) Studies on the constituents of domestic *Lycopodium* genus plants XIII. On the constituents of *Lycopodium cernuum* and *Lycopodium inuddatum* L. *Yakugaku Zasshi*, 91, 980-986.
- INUBUSHI, Y., TSUDA, Y., ISHII, H., HOSOKAWA, N., & SANO, T. (1962) *Yakugaku Zasshi*, 82, 1339.
- ITURBE-ORMAETXE, I., HARALAMPIDIS, K., PAPADOPOULOU, K., & OSBOURN, A.E. (2003) Molecular cloning and characterization of triterpene synthases from *Medicago truncatula* and *Lotus japonicus*. *Plant Molecular Biology*, 51, 731-743.
- IVIMEY-COOK, R.B. (1968) Investigations into the phonetic relationships between species of *Ononis* L. *Watsonia*, 7, (1), 1-23.
- JACOB, J., DISNAR, J.R., BOUSSAFIR, M., ALBUQUERQUE, A.L.S.A., SIFADDINE, A., & TURCQ, B. (2004) Onocerane I testimonies to dry climatic events during the Quaternary in the Tropics. *Organic Geochemistry*, 35, 289-297.
- JARIC, S., POPVIC, Z., JOCIC, M.M., DJURDJEVIC, L., MIJATOVIC, M., KARIDIZIC, B., MITRVIC, M., & PAVLOVIC, P. (2007) An ethnobotanical study on the usage of wild medicinal herbs from Kopaonik Mountain (Central Serbia). *Journal of Ethnopharmacology*, 111, 160-175.
- KAJIKAWA, M., YAMATO, K.T., FUKUZAWA, H., SAKAI, Y., UCHINDA, H., & OHYAMA, K. (2005) Cloning and characterization of cDNA encoding  $\beta$ -amyrin synthase from petroleum plant *Euphorbia tirucalli* L. *Phytochemistry*, 66, 1759-1766.
- KARLOVSKY, P. (2008) *Secondary Metabolites in Soil Ecology Soil Biology*. Springer-Verlag Berlin, Heidelberg.
- KLODA, J.M., DEAN, P.D.G., MACDONALD, D.W., & MAYES, S. (2004) Isolation and characterization of microsatellite loci in *Ononis repens*, Leguminosae. *Molecular Ecology Notes*, 4, 596-598.
- KLODA, J.M., DEAN, P.D.G., MACDONALD, D.W., & MAYES, S. (2008) Using principle component analysis to compare genetic diversity

across polyploidy levels within plant complexes: an example from British Restharrowes (*Ononis spinosa* and *Ononis repens*). *Heredity*, 100, 253-260.

KOLESNIKOVA, M.D., XIONG, Q., LODERIO, S., HUA, L., & MATSUDA, S.P.T. (2006) Lanosterol biosynthesis in plants. *Archives of Biochemistry and biophysics*, 447, 87-95.

KUSHIRO, T., SHIBUYA, M., MASUDA, K., & EBIZUKA, Y. (1998)  $\beta$ -amyrin synthase; Cloning of oxidosqualene cyclase that catalyzes the formation of the most popular triterpene among higher plants. *European Journal of Biochemistry*, 256, 238-244.

KUSHIRO, T., SHIBUYA, M., MASUDA, K., & EBIZUKA, Y. (2000) Mutational studies on triterpene synthases: engineering lupeol synthase into  $\beta$ -amyrin synthase. *American Chemical Society*, 122, 6816-6824.

LANGE, M.B., & GHASSEMIAN, M. (2003) Genome organization in *Arabidopsis thaliana*: a survey for genes involved in isoprenoid and chlorophyll metabolism. *Plant Molecular Biology*, 51, 925-948.

LEE, M.H., JEONG, J.H., SEO, J.W., SHIN, C.G., KIM, Y.S., IN, J.G., YANG, D.C., YI, J.S., & CHOI, Y.E. (2004) Enhanced triterpene and phytosterol biosynthesis in *Panax ginseng* over expressing squalene synthase gene. *Plant Cell Physiology*, 45:8:976-984.

LEI, Z., HUHMANN, D.V., SUMNER, L.W. (2011) Mass spectrometry strategies in metabolomics. *Journal of Biological Chemistry*, 286:29, 25435-25442.

LEWIS, D.F.V., WATSON, E., & LAKE, B.G. (1997) Evolution of the cytochrome P450 superfamily: sequence alignments and pharmacogenetics. *Mutation Research*, 410(3), 245-270.

LIANG, Y., & ZHAO, S. (2007) Progress in understanding of ginsenoside biosynthesis. *Plant Biology*, 10(4), 415-421.

LIEBEZEIT, G. (2007) Ethnobotany and phytochemistry of plants dominant in salt marshes of the Lower Saxonian Wadden Sea, southern North Sea. *Marine Biodiversity*, 38(1), 1-30.

LIU, W.M., MEI, R., DI, X., RYDER, T.B., HUBBELL, E., DEE, S., WEBSTER, T.A., HARRINGTON, A., HO, M.H., BAID, J., & SMEEKENS, S.P. (2002) Analysis of high density expression microarrays with signed-rank call algorithms. *Bioinformatics*, 18, 1593–1599.

LOCKHART, D.J., DONG, H., BYRNE, M.C., FOLLETTIE, M.T., GALLO, M.V., CHEE, M.S., MITTMANN, M., WANG, C., LODEIRO, S., SCHULZ-GASCH, T., & MATSUDA, S.P.T. (2005) *Journal of the American Chemical Society*, 127, 14132-14133.

LOCKHART, D. J., DONG, H., BYRNE, M. C., FOLLETTIE, M. T., GALLO, M. V., CHEE, M. S., MITTMANN, M., WANG, C.,



- KOBAYASHI, M., HORTON, H. AND BROWN, E. L. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* 14, 1675-1680.
- MAHASNEH, A.M., & EL-OQLAH, A.A. (1999) Antimicrobial activity of extracts of herbal plants used in the traditional medicine of Jordan. *Journal of Ethnopharmacology*, 64, 271-276.
- MARMUR, J., & LANE, D. (1960) Strand separation and specific recombination in deoxyribonucleic acids: biological studies. *Proceedings of the National Academy of Sciences*, 46, 453-461.
- MATSUDA, K., SHIOJIMA, K., & AGETA, H. (1989) Fern constituents: four new onoceradienes isolated from *Lemmaphyllum microphyllum* Presl. *Chemical and Pharmaceutical Bulletin*, 3, 263-265.
- MEUNIER, B., DE VISSER, S.P., & SHAIK, S. (2004) Mechanisms of Oxidation Reactions Catalyzed by cytochrome P450 Enzymes. *Chemical Reviews*, 104(9), 3947-3980.
- MI, Y., SCHREIBER, J.V., & COREY, E.J. (2002) Total synthesis of (+)- $\alpha$ -onocerin in Four steps via four-component coupling and tetracyclization steps. *Journal of American Chemical Society*, 124, 11290-11291.
- MIGAHID, & M.M. (2003) Effect of salinity shock on some desert species native to the northern part of Egypt. *Journal of Arid Environment*, 53, 155-167.
- MORISSET, P. (1967) Cytological and Taxonomic studies in *Ononis spinosa* and *O. repens* and related species. *Thesis University of Cambridge*.
- MORISSET, P. (1978) Chromosome numbers in *Ononis* L. Series *Vulgares* Īrj. *Watsonia*, 12, 145-153.
- MORITA, M., SHIBUYA, M., KUSHIRO, T., MASUDA, K., & EBIZUKA, Y. (2008) Molecular cloning and functional expression of triterpene synthases from pea (*Pisum sativum*). *European Journal of Biochemistry*, 267, 3453-3460.
- MORRIS, M.G. (1990) Orthocerous weevils. *Royal Entomological Society of London*, 5(16), 108.
- MUELLER, L. A., PEIFEN, Z., Rhee, S.Y. (2003) A Biochemical Pathway Database for *Arabidopsis*, *Plant Physiology*, 132, 453-460.
- MURPHY, S.T., RITCHIE, E., & TALYOR, W.C. (1974) Constituents of *Melicope perspicuinervia* (Rutaceae). The structures of halfordinine and Melinervin a New Flavone. *Australian Journal of Chemistry*, 27, 187-94.
- NELSON, J.A., STECKBECK, S.R., & SPENCER, T.A. (1981) Biosynthesis of 24,25-epoxycholesterol from squalene 2,3;22,23-dioxide. *Journal of Biological Chemistry*, 256, 1067-1068.

- NISHIZAWA, M., NISHIDE, H., KOSELA, S., & HAYASHI, Y. (1983) Structure of Lansiosides: biologically active new triterpene glycosides from *Lansium domesticum*. *Journal of Organic Chemistry*, 48, 4462-4466.
- ORHAN, I., TERZIOGLU, S., & SENER, B. (2003)  $\alpha$ -onocerin: An acetylcholinesterase inhibitor from *Lycopodium clavatum*. *Planta Medica*, 69, 265-267.
- PAPADOPOULOU, K., MELTON, R.E., LEGGETT, M., DANIELS, M.J., & OSBOURN, A.E. (1999) Compromised disease resistance in saponin - deficient plants. *Proceedings of the National Academy of Sciences. USA*, 96, 12923-12928.
- PAULI, G.F. (2000) Comprehensive spectroscopic investigation of  $\alpha$ -onocerin. *Planta Medica*, 66, 299-302.
- PAULI, G.F., & FROHLICH, R. (2000) Chiral key positions in Uzara steroids. *Phytochemical Analysis*, 11, 79-89.
- PHILLIPS, D.R., RASBERY, J.M., BARTEL, B., & MATSUDA, S.P.T. (2006) Biosynthetic diversity in plant triterpene cyclisation. *Current Opinion in Plant Biology*, 9, 305-314.
- QI, X., BAKHT, S., LEGGETT, M., MAXWELL, C., MELTON, R., & OSBOURN, A. (2004) A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants. *Proceedings of the National Academy of Sciences*, 101, 8233-8238.
- RASBERY, J.M., HUI, S., LECLAIR, R.J., NORMAN, M., MATSUDA, S.P.T., & BARTEL, B. (2007) *Arabidopsis thaliana* Squalene Epoxidase 1 Is Essential for Root and Seed Development. *Journal of Biological Chemistry*, 10, 1074.
- ROLLINGER, J.M., EWELT, J., SERGER, C., STURM, S., ELLMERER, E.P., & STUPPNER, H. (2005) New insights into the acetylcholinesterase inhibitory activity of *Lycopodium clavatum*. *Planta Medica*, 71, 1040-1043.
- ROWAN, M.G. (1971) Some studies on the occurrence and biosynthesis of alpha-onocerin. *Thesis University of Liverpool*.
- ROWAN, M.G., & DEAN, P.G. (1972a) Properties of squalene-2(3),22(23)-diepoxide- $\alpha$ -onocerin cyclase from *Ononis spinosa* root. *Phytochemistry*, 11, 3111-3118.
- ROWAN, M.G., & DEAN, P.D.G (1972b).  $\alpha$ -Onocerin and sterol content of 12 species of *Ononis*. *Phytochemistry*, 11, 3263-3265.
- RUZICKA, L., ESCHENMOSER, A., & HEUSSER, H. (1953) *Experientia*, 9, 357.
- RUZICKA, L. (1959) *Proceedings of the Chemical Society of London*, 341-360.

- SAWAI, S., SHINDO, T., SATO, S., KANEKO, T., TABATA, S., AYABE, S., & AOKI, T. (2006) Functional and structural analysis of genes encoding oxidosqualene cyclases of *Lotus japonicus*. *Plant Science*, 170, 247-257.
- SCHAFFNER, K., VITERBO, R., ARIGONI, & D., JEGER. (1956) *Helvetica Chimica Acta*, 39, 174.
- SCHENA, M., SHALON, D., HELLER, R., CHAI, A., BROWN P. O., & DAVIS, R.W. (1996) Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proceedings of the National Academy of Sciences*, 93(20), 10614-10619.
- SCHREIBER, F., PICK, K., ERPENBECK, D., WORHEIDE, G., & MORGENSTERN, B. (2009) OrthoSelect: a protocol for selecting orthologous groups in phylogenomics *BMC Bioinformatics*, 10, 219.
- SEGURA, M.J.R., MEYER, M.M., ZAKRAJSEK, M., & MATSUDA, S.P.T. (2000) *Arabidopsis thaliana* LUP1 converts oxidosqualene to multiple triterpene alcohols and a triterpene diol. *American Chemical Society*, 2(15), 2257-2259.
- SHALLARI, S., SCHWARTZ, C., HASKO, A., & MOREL, J.L. (1998) Heavy metals in soils and plants of serpentine and industrial sites of Albania. *The Science of the Total Environment*, 209, 133-142.
- SHIBATA, N., ARITA, M., MISAKI, Y., DOHMAE, N., TAKIO, K., ONO, T., INONUE, K., & ARIA, H. (2001) Supernatant protein factor, which stimulates the conversion of squalene to lanosterol, is a cytosolic squalene transfer protein and enhances cholesterol biosynthesis. *Proceedings of the National Academy of Science*, 98, 2244-2249.
- SHIBUYA, M., XIANG, T., KATSUBE, Y., OTSUKA, M., ZHANG, H., & EBIZUKA, Y. (2007) Origin of structural diversity in natural triterpenes: Direct synthesis of *seco*-triterpene skeletons by Oxidosqualene cyclase. *Journal of the American Chemical Society*, 129(5), 1450-1455.
- SHIBUYA, M., ZHANG, H., ENDO, A., SHISHIKURA, K., KUSHIRO, T., & EBIZUKA, Y. (1999) Two branches of the lupeol synthase gene in the molecular evolution of plant oxidosqualene cyclises. *Federation of European Biochemical Societies Letters*, 266, 302-307.
- SHUAN, H., SEGURA, M.J.R., WILSON, W.K., LODEIRO, S., & MATSUDA, P.T. (2005) Enzymatic cyclization of dioxidosqualene to heterocyclic triterpenes. *American Chemical Society*, 127(51), 18008-18009.
- SIGMA-ALDRICH CO. TECHNICAL INFORMATION (1997) Bulletin 909A, *In Guide to derivitisation reagents for GC*. Supelco Harrison road Bellefonte.

- SOUTHERN, EDWIN MELLOR (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology*, 98(3), 503–517.
- SPENCER, K.A. (1972) Diptera Agromyzidae. *Royal Entomological society of London*, 10(5), 136.
- SPIKETT, N., SMIRNOFF, R., & RATCLIFFE, G. (1993) An in vivo nuclear magnetic resonance investigation of ion transport in maize (*Zea mays*) and *Spartina anglica* roots during exposure to high salt concentrations. *Plant Physiology*, 102(2), 629-638.
- STAFFORD, H.A. (2008) Roles of flavonoids in symbiotic and defence functions in legume roots. *The Botanical Review*, 63, 27-39.
- STERKY, F., BHALERAO, R.R., UNNEBERG, P., SEGERMAN, B., NILSSON, & P., BRUNNER, A.M. (2005) A Populus EST resource for plant functional genomics. *Proceeding of the National Academy of Sciences*. 101, 13951-13956.
- STORK, G., MEISELS, A., & DAVIES J.E. (1963) Total synthesis of polycyclic triterpenes: total synthesis of (+) –  $\alpha$ -onocerin. *Journal of the American Chemical Society*, 85, 3419.
- SUMNER, L.W., MENDES, P., & DIXON, R.A. (2003) Plant metabolomics: large scale phytochemistry in the functional genomics era. *Phytochemistry*, 62, 817-36.
- SUZUKI, H., ACHNINE, L., XU, R., MATSUDA, S.P.T., & DIXON, R.A. (2002) A genomics approach to the early stages of triterpene saponin biosynthesis in *Medicago truncatula*. *The Plant Journal*, 32, 1033-1048.
- TAKASAKI, M., KONOSHIMA, T., TOKUDA, H., MASUDA, K., ARAI, Y., SHIOJIMA, K., & AGETA, H. (1999) *Biological and Pharmacological Bulletin*, 45, 2016-2023.
- TOHGE, T., NISHIYAMA, Y., HIRAI, MY., YANO, M., NAKAJIMA, J., AWAZUHARA, M., INOUE, E., TAKAHASHI, H., GOODENOWE, D.B., KITAYAMA, M., NOJI, M., YAMAZAKI, M., & SAITO, K. (2005) Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. *Plant Journal*, 42, 218-35.
- TSUDA, Y., FUJIMOTO, T., & KIMPARI K. (1970) *Journal of the Chemical Society*, 9, 261.
- VAN TAMELEN, E.E., & HOPLA, R.E. (1979) Generation of the onocerin system by lanosterol 2,3-oxidosqualene cyclase-implications for the cyclization process. *American Chemical Society*, 101(20), 6112-6114.
- VON HEMMELMAYR, F. (1907) *Montashefte*, 27, 181.

- WANG, X., KREPS, J.A., & KAY, S.A. (2000) Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science*, 290, 2110-2113.
- WANG, Z, GERSTEIN, M., & SNYDER, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews. Genetics*, 10(1), 57-63.
- WINK, M. (1999). *Functions of plant secondary metabolites and their exploitation in biotechnology*. Sheffield, UK: Sheffield Academic Press.
- XIONG, Q., WILSON, W.K., & MATSUDA, S.P.T. (2006) An *Arabidopsis* Oxidosqualene Cyclase Catalyzes Iridal Skeleton Formation by Grob Fragmentation. *Angewandte Chemical International Edition*, 45, 1285-1288.
- XU, R., FAZIO, G.C., & MATSUDA, S.P.T. (2004) On the origins of triterpenoid skeletal diversity *Phytochemistry*, 65(3), 261-291.
- YODER, R.A., JOHNSTON, J.N. (2005) A case study in biometric total synthesis: polyolefin carbocyclizations to triterpenes and steroids. *Natural product synthesis thematic issue*, 105,730.
- YOLMAZ, B.S., HANEFI, O., GULCIN, S.I., SERDAR, U., IRFAN, B., & ENDER, E. (2006) Analgesic and hepatotoxic effects of *Ononis spinosa* L. *Phytotherapy Research*, 20(6), 500-503.
- ZHANG, Z.Z., ELSOHL, H.N., & JACOB, M.R. (2002) Natural products inhibiting *Candida albicans* secreted aspartic proteases from *Lycopodium cernuum*. *Journal of Natural Products*, 65, 979-985.
- ZHONG, J.J & YUE C.J. (2005) Secondary Metabolite Heterogeneity and Its Manipulation. *Advances in Biochemical Engineering/Biotechnology*, 100, 53-88.
- ZIMMERMAN, J. (1938) *Helvetica Chimica Acta* 21, 853.
- ZIMMERMAN, P., HIRSCH-HOFFMANN, M., HENNIG, L., & WILHELM, G. (2004) GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. *Plant Physiology*, 136, 2621-2632.

## APPENDIX

### APPENDIX I: MATERIALS AND METHODS USED IN THE THESIS

#### Materials

Chemical	Supplier
Agar	Fisher Scientific, UK
Agarose Electrophoresis grade	Invitrogen, UK
Agarose type PGP	Park Scientific Ltd, UK
$\beta$ -amyrin	Dr P.D.G Dean (gift)
Ampicillin sodium salt	Sigma-Aldrich Co. Ltd., UK
Ascorbate	Fisher Scientific, UK
Boric acid	Melford, UK
Bromophenol blue	Sigma-Aldrich Co. Ltd., UK
Chloral hydrate	Sigma-Aldrich Co. Ltd., UK
Chloroform	Fisher Scientific, UK
Chromatographic alumina	Acros organics
Cyclohexane	Fisher Scientific, UK
Cystiene	Oxoid, UK
DH5 $\alpha$ Chemically competent cells	University of Nottingham, UK
Dichloromethane	Fisher Scientific, UK
Diethyl ether	Fisher Scientific, UK
Dimethyl formamide (DMF)	Fisher Scientific, UK
2,3:22,23-dioxidosqualene	Echlon
EDTA	Sigma-Aldrich Co. Ltd., UK
Ethanol Absolute	Fisher Scientific, UK
Ficoll Type 4000	Fisher Scientific, UK
Filter 0.22 $\mu$ m	Anachem, UK
Fuchsin acid	Fisher Scientific Co. Ltd, UK
Glucose	Fisher Scientific, UK
Helium	BOC?
Hexane	Fisher Scientific, UK
Hydrochloric acid	Fisher Scientific, UK
IPTG (Isopropylthio-b-D-galactoside)	Apollo
John Innes compost	Scott's, UK
Kanamycin sulfate	Sigma-Aldrich Co. Ltd., UK
Liquid Nitrogen	BOC
Magnesium Chloride	Fisher Scientific, UK
Magnesium sulphate	Fisher Scientific, UK
Methanol	Fisher Scientific, UK
MS (Murashige and Skoog medium)	Duchefa Biochemie,
Nitrogen	BOC
Octacosane	Acros organics, UK
Oligonucleotides	MWG Eurofins, UK
$\alpha$ -onocerin	Echlon Biosciences Inc., USA
2,3-oxidosqualene	Echlon Biosciences Inc., USA
Perlite	Scott's, UK
Continued...	
Phenol	Sigma-Aldrich Co. Ltd., UK

Polyvinylpyrrolidine	Fisher Scientific, UK
Potassium Chloride	Fisher Scientific, UK
Potassium Iodide	Fisher Scientific, UK
Potassium hydroxide	Fisher Scientific, UK
Potassium phosphate	Fisher Scientific, UK
Propan-2-ol	Fisher Scientific, UK
Sodium Chloride	Fisher Scientific, UK
Sodium hypochlorate	Parazone,UK
Squalene	Echlon
Sucrose	Fisher Scientific, UK
Triacontane	Acros Organics, UK
Tris base	Melford, UK
Trizol	Sigma-Aldrich Co. Ltd., UK
Tryptone	Oxoid, UK
Tween 80	Acros Organics, UK
X-GAL (5-Bromo-4-chloro-3-indolyl $\beta$ -	Apollo, UK
Yeast Extract granulate	MERCK, Germany

## Equipment

Equipment/Apparatus	Company
37°C incubator	Gallenkamp,UK
50 ml tube centrifuge; Centaur 2	Shelton Technical Ltd., UK
BioRad MicroPulser™	Bio-Rad Lab, UK
Cold Centrifuge; Centrifuge 5810R	Eppendorf, Germany
Digital camera with Stereo microscope;	Leica Microsystem Ltd., UK
GC/MS 8000 Series (MD800)	Fisons
Gel-electrophoresis power supply;	Biorad lab, UK
Gel-electrophoresis tanks; Sub-	Biorad lab, UK
Gene Amp PCR system 9700(PCR)	Applied Biosystems
Hotblock- DRI-Block DB-3	Techne, UK
Large Autoclave 400/500L	Boxer Lab Equipment Ltd. UK
Large orbital incubator; Gallenkamp,	Sanyo-biomedical, UK
LightCycler® 480 instrument	Roche Applied Science, UK
Microcentaur	MSE
Microcentrifuge; Sigma 1-15, 10115	Sigma, Germany
MX3005p real-time PCR machine	Stratagene
Nanodrop (ND-100)	Thermo Scientific
Omega media prestige medical	Prestige™Medical, England
pH meter; SevenEasy pH	Mettler Toledo, UK
Spectrophotometer; SP6-500 UV	Pye Unicam Ltd., England
Stereo microscope; Leica MZ6	Leica Microsystem Ltd., UK
Thermo Fisher DSQII Mass	ThermoFisher , UK
Thermo Hybaid PCR machine	Thermo Electron corporation
UV image Capture-Gel Doc 2000	Biorad lab, UK
Video copy processor p91	Mitsubishi
Vortex machine; Genie-2	Scientific industries Inc., USA
Water bath (Grant JB series)	Grant Instruments (Cambridge)

Weight 2 digits; Precisa Junior 500C and	Precisa Instrument Ltd.,
Weight 4 digits; Precisa XB120A	Precisa Instrument Ltd..

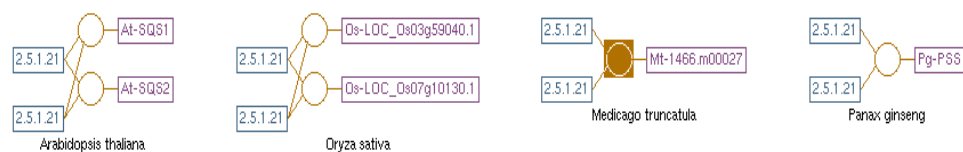
# **Software used within the thesis**

<b>Organization</b>	<b>Web link</b>
The National Center for	<a href="http://www.ncbi.nlm.nih.gov">http://www.ncbi.nlm.nih.gov</a>
European Bioinformatics	<a href="http://www.ebi.ac.uk">http://www.ebi.ac.uk</a>
The Institute for Genomic	<a href="http://www.tigr.org/plantProjects.shtml">http://www.tigr.org/plantProjects.shtml</a>
The Arabidopsis Information	<a href="http://www.arabidopsis.org">http://www.arabidopsis.org</a>
PLEXdb (Plant expression	<a href="http://www.plexdb.org">http://www.plexdb.org</a>
UniProt Protein database.	<a href="http://www.uniprot.org/help/uniprotkb">http://www.uniprot.org/help/uniprotkb</a>

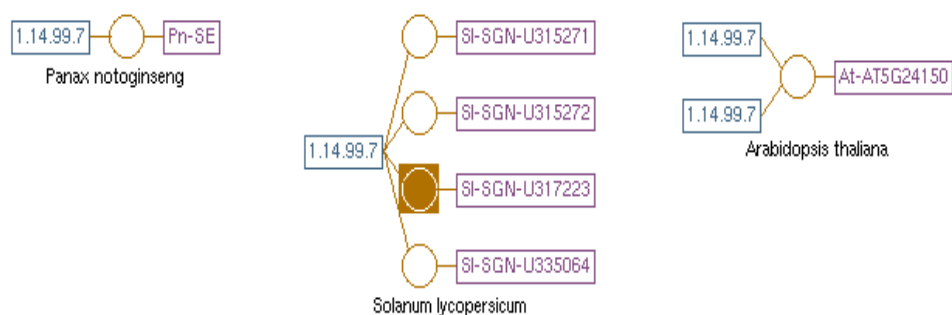
<b>Software</b>	<b>Website link</b>	<b>Application</b>
Accelrys Gene Version	<a href="http://www.accelrys">http://www.accelrys</a>	Nucleic and amino
EPI Suit <sup>TM</sup> vs 4.0	<a href="http://www.epa.gov/op">http://www.epa.gov/op</a>	Chemical properties
Geldoc		UV gel analysis
GeneSpring GX	<a href="http://www.chem.agile">http://www.chem.agile</a>	Transcriptional
GenStat for Windows	<a href="http://www.vsni.co.uk/">http://www.vsni.co.uk/</a>	RMA analysis of
Leica IM50 version 4.0,	<a href="http://www.chem.agile">http://www.chem.agile</a>	Viewing root and
Mass lab		GC/MS programs
Mass Lynx 4.1		GC/MS, analysis
NIST library:		Mass spectral library
PIGEONS	Sean Mayes the	gDNA hybridisation
SciFinder Scholar	<a href="http://www.cas.org/pro">http://www.cas.org/pro</a>	Chemical parameters:
Swiss Model Server and	<a href="http://swissmodel.exp">http://swissmodel.exp</a>	Protein modelling
Xcalibur version 1.4	<a href="http://www.thermo.co">http://www.thermo.co</a>	develop; program



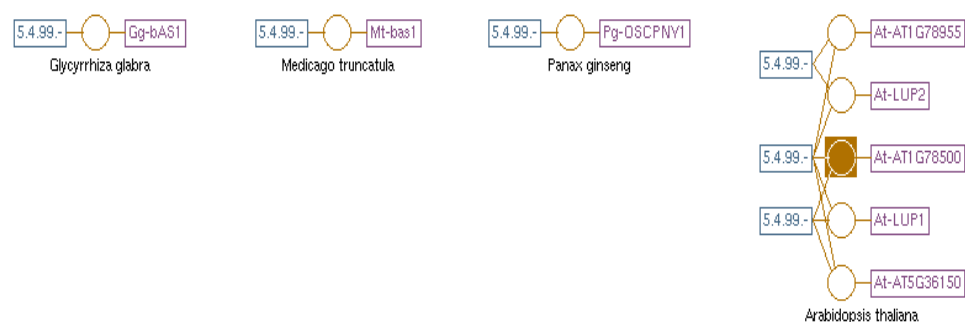
## APPENDIX II: ARACYC TRITERPENE PATHWAY OUTPUT



Output of squalene synthases from the Plant Metabolic Network (PMN) viewer (<http://pmn.plantcyc.org>).

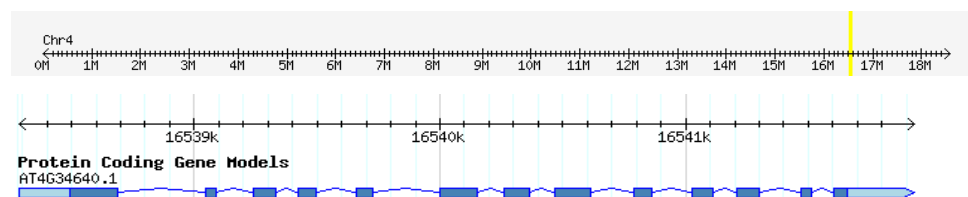


A summary showing the output of squalene epoxidase from the Plant Metabolic Network (PMN) viewer (<http://pmn.plantcyc.org>).



Output of  $\beta$ -amyirin synthases from the Plant Metabolic Network (PMN) viewer (<http://pmn.plantcyc.org>).

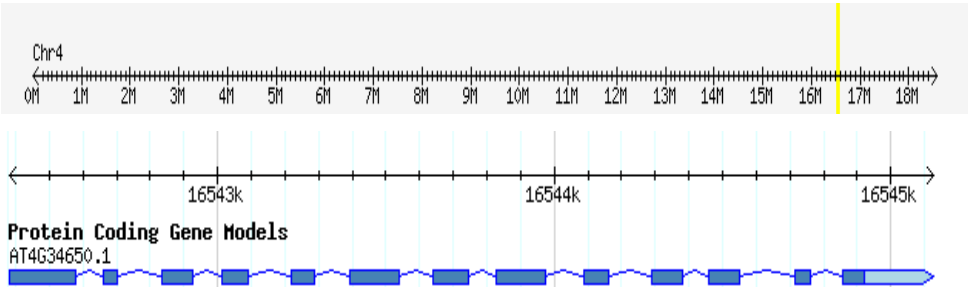
# APPENDIX III GENE MODELS OF TRITERPENE BIOSYNTHESIS IN *A. THALIANA* Squalene synthase



Squalene synthase 1 (SQS1): At4g34640 gene structure and position on chromosome 4. Number of exons retrieved: 13

#	Exon Contents	5' End	3' End	+ strand	- strand	Size (bp)	Beginning of codon:
1	5' UTR + ORF	16,538,287	16,538,688	1,3,5	2,4,6	401	
Intron				7	8		
2	ORF	16,539,047	16,539,089			42	
Intron							
3	ORF	16,539,239	16,539,328			89	
Intron							
4	ORF	16,539,420	16,539,495			75	
Intron							
5	ORF	16,539,657	16,539,726			69	
Intron				9,11	10,12		
6	ORF	16,540,001	16,540,150	13	14	149	
Intron							
7	ORF	16,540,259	16,540,363			104	
Intron							
8	ORF	16,540,466	16,540,612	15,17	16	146	
Intron				17,18			
9	ORF	16,540,788	16,540,863		19	75	
Intron							
10	ORF	16,541,024	16,541,112	20	21	88	
Intron							
11	ORF	16,541,210	16,541,301	22	23	91	
Intron							
12	ORF	16,541,468	16,541,510			42	
Intron							
13	ORF + 3' UTR	16,541,604	16,541,931	24,25	26	327	
Within 3' UTR							

Squalene synthase 2

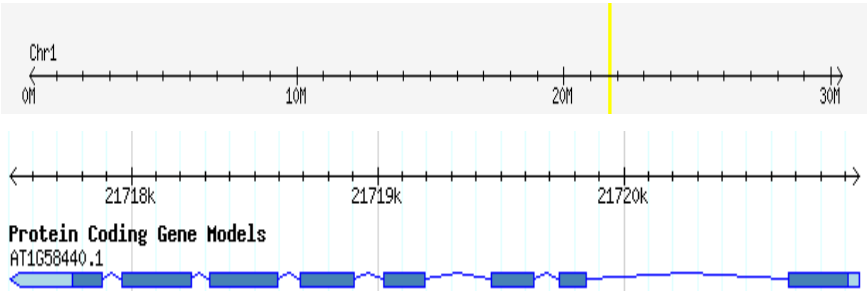


Squalene synthase 2

Number of exons retrieved: 13

#	Exon Contents	5' End	3' End	+ strand	- strand	Size (bp)	Beginning of start codon: bp
1	ORF	16,542,382	16,542,576			194	16 542 382
Intron							
2	ORF	16,542,661	16,542,703			42	
Intron							
3	ORF	16,542,837	16,542,926	1	2	89	
Intron							
4	ORF	16,543,014	16,543,089			75	
Intron				3	4		
5	ORF	16,543,217	16,543,286			69	
Intron							
6	ORF	16,543,391	16,543,540			149	
Intron							
7	ORF	16,543,642	16,543,746			104	
Intron				5	6		
8	ORF	16,543,827	16,543,973	7,9	8	146	
Intron				9,10			
9	ORF	16,544,087	16,544,162		11	75	
Intron							
10	ORF	16,544,292	16,544,380	12	13	88	
Intron							
11	ORF	16,544,457	16,544,548	14	15	91	
Intron				16,18	17,19		
12	ORF	16,544,716	16,544,758			42	
Intron							
13	ORF + 3' UTR	16,544,855	16,545,126	20,21	22	271	
Within 3' UTR				23			

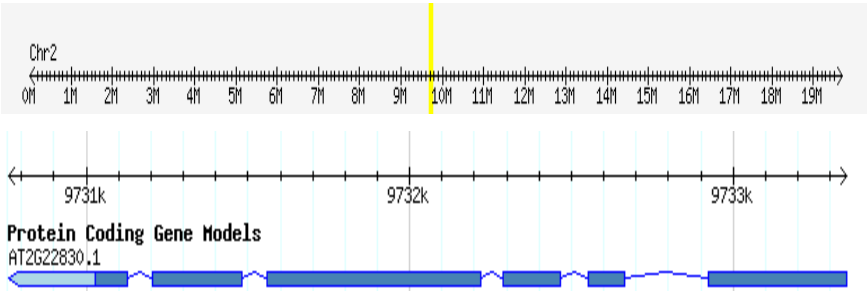
Squalene epoxidase 1



Squalene epoxidase 1 Number of exons retrieved: 8

#	Exon Contents	5' End	3' End	+ strand	- strand	Size (bp)	Beginning of start codon: 21,720,912 bp End of stop codon: 21,717,758 bp
1	5' UTR + ORF	21,720,955	21,720,669	16,18,20,22	17,19,21,23	286	
Intron							
2	ORF	21,719,849	21,719,738	10	11	111	
Intron							
3	ORF	21,719,636	21,719,462			174	
Intron							
4	ORF	21,719,194	21,719,026			168	
Intron							
5	ORF	21,718,905	21,718,687	6	7	218	
Intron							
6	ORF	21,718,594	21,718,319	4	5	275	
Intron							
7	ORF	21,718,242	21,717,964			278	
Intron							
8	ORF + 3' UTR	21,717,881	21,717,507			374	
Within 3' UTR							

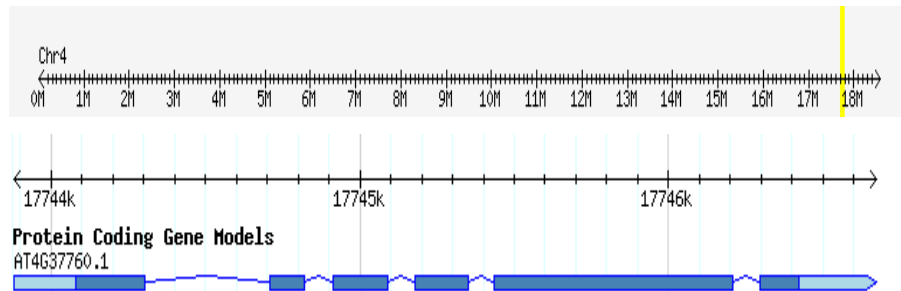
Squalene epoxidase 2



Number of exons retrieved: 6

#	Exon Contents	5' End	3' End	+ strand	- strand	Size (bp)Beginning of start codon: 9,733,350 bp End of stop codon: 9,731,026 bp
1	ORF	9,733,350	9,732,922	17,19,21	18,20,22	428
Intron						
2	ORF	9,732,662	9,732,551			111
Intron						
3	ORF	9,732,463	9,732,289			174
Intron						
4	ORF	9,732,220	9,731,557	9,11,13	10,12,14	663
Intron						
5	ORF	9,731,481	9,731,203	7	8	278
Intron						
6	ORF + 3' UTR	9,731,125	9,730,768	2	3,4	357
Within 3' UTR						

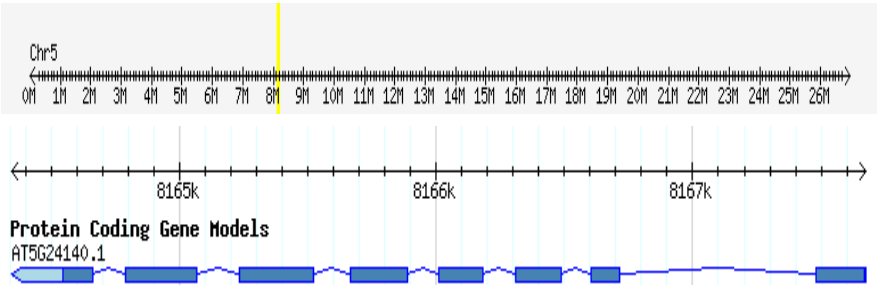
### Squalene epoxidase 3



Number of exons retrieved: 6

#	Exon Contents	5' End	3' End	+ strand	- strand	Size (bp)Beginning of start codon:
1	5' UTR + ORF	17,743,881	17,744,303	1,3	2,4	422
Intron						
2	ORF	17,744,708	17,744,819	7	8	111
Intron						
3	ORF	17,744,914	17,745,088	9	10	174
Intron						
4	ORF	17,745,180	17,745,348	11	12	168
Intron						
5	ORF	17,745,435	17,746,208	17,19,21,23	18,20,22,24	773
Intron						
6	ORF + 3' UTR	17,746,298	17,746,672	25,27,28	26,29	374
Within 3' UTR						

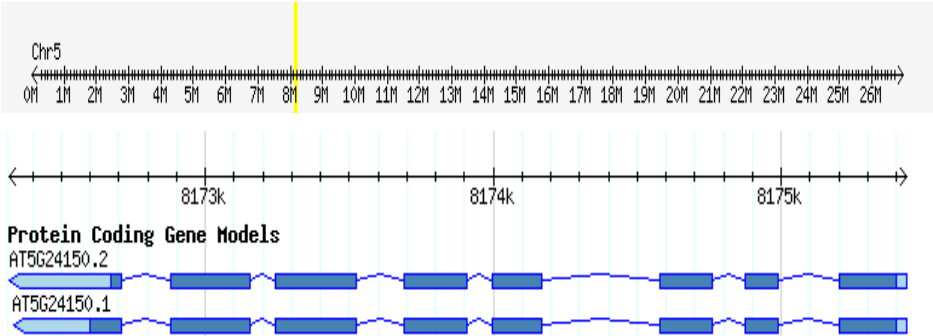
Squalene epoxidase 4



Number of exons retrieved: 8

#	Exon Contents	5' End	3' End	+ strand	- strand	Size (bp)
1	5' UTR + ORF	8,167,676	8,167,481			195
Intron						
2	ORF	8,166,719	8,166,608			111
Intron						
3	ORF	8,166,489	8,166,312			177
Intron						
4	ORF	8,166,183	8,166,015			168
Intron						
5	ORF	8,165,890	8,165,672	9	10	218
Intron						
6	ORF	8,165,522	8,165,235	5	6	287
Intron						
7	ORF	8,165,069	8,164,791			278
Intron						
8	ORF + 3' UTR	8,164,663	8,164,359	3	4	304
Within 3' UTR						

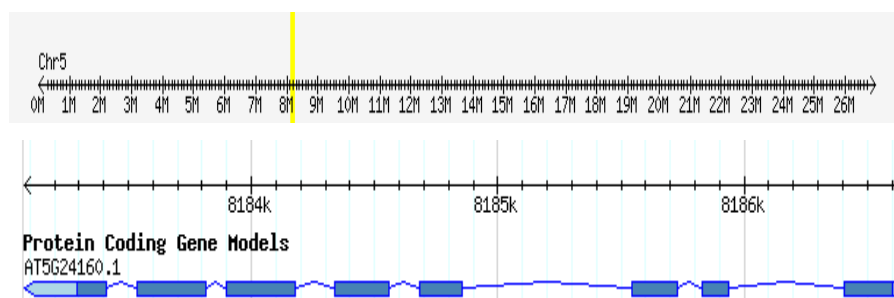
Squalene epoxidase 5



Number of exons retrieved: 8

#	Exon Contents	5' End	3' End	+ strand	- strand	Size (bp)	Beginning of start codon: 8,175,399 bp
1	5' UTR + ORF	8,175,434	8,175,201			233	
Intron							
2	ORF	8,174,985	8,174,874			111	
Intron							
3	ORF	8,174,760	8,174,580	9,11	10,12	180	
Intron							
4	ORF	8,174,166	8,173,998	5	6	168	
Intron							
5	ORF	8,173,907	8,173,689			218	
Intron							
6	ORF	8,173,522	8,173,244			278	
Intron							
7	ORF	8,173,156	8,172,878	1	2	278	
Intron							
8	ORF + 3' UTR	8,172,710	8,172,335			375	
Within 3' UTR							

## Squalene epoxidase 6

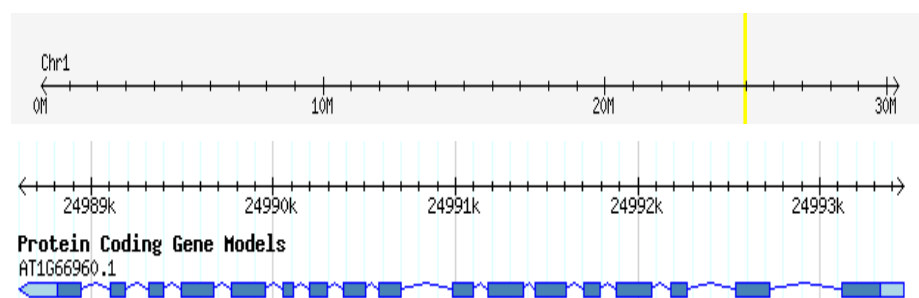


Number of exons retrieved: 8

#	Exon Contents	5' End	3' End	+ strand	- strand	Size (bp)	Beginning of start codon: 8,186,605 bp
1	5' UTR + ORF	8,186,643	8,186,407	8	9	236	
Intron							
2	ORF	8,185,938	8,185,827	6	7	111	
Intron							
3	ORF	8,185,725	8,185,545	4	5	180	
Intron							
4	ORF	8,184,855	8,184,684			171	
Intron							
5	ORF	8,184,554	8,184,336			218	
Intron							
6	ORF	8,184,179	8,183,901			278	
Intron							
7	ORF	8,183,816	8,183,538	2	3	278	
Intron							
8	ORF + 3' UTR	8,183,407	8,183,078			329	
Within 3' UTR							

# Oxidosqualene cyclases

## Lupeol synthase

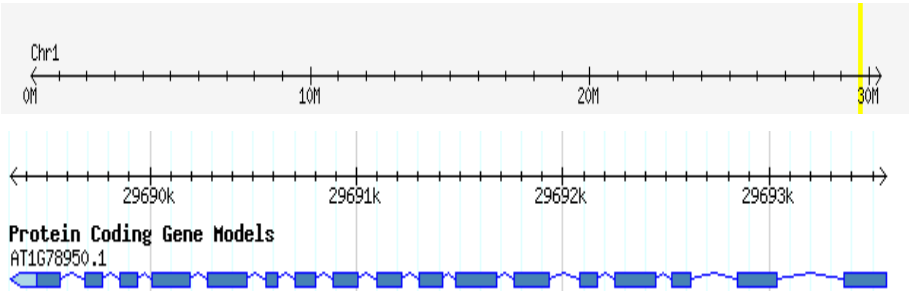


Number of exons retrieved: 17

#	Exon Contents	5' End	3' End	+ strand	- strand	Size (bp)	Beginning of start codon: 24,993,328 bp
1	5' UTR + ORF	24,993,460	24,993,124	35,37	36,38	336	
Intron							
2	ORF	24,992,724	24,992,539			185	
Intron							
3	ORF	24,992,267	24,992,178			89	
Intron							
4	ORF	24,992,076	24,991,879			197	
Intron							
5	ORF	24,991,790	24,991,706			84	
Intron							
6	ORF	24,991,604	24,991,438			166	
Intron							
7	ORF	24,991,372	24,991,181			191	
Intron							
8	ORF	24,991,097	24,990,984			113	
Intron							
9	ORF	24,990,701	24,990,582	25	26	119	
Intron							
10	ORF	24,990,506	24,990,384	21,23	22,24	122	
Intron							
11	ORF	24,990,297	24,990,199			98	
Intron							
12	ORF	24,990,112	24,990,056			56	
Intron							
13	ORF	24,989,956	24,989,766	15,17	16,18	190	
Intron							
14	ORF	24,989,669	24,989,492			177	
Intron							
15	ORF	24,989,396	24,989,316	13	14	80	
Intron							
16	ORF	24,989,187	24,989,104	11	12	83	
Intron							
17	ORF + 3' UTR	24,988,940	24,988,604	1,4,7	2,3,5,6,8	336	
Within 3' UTR							



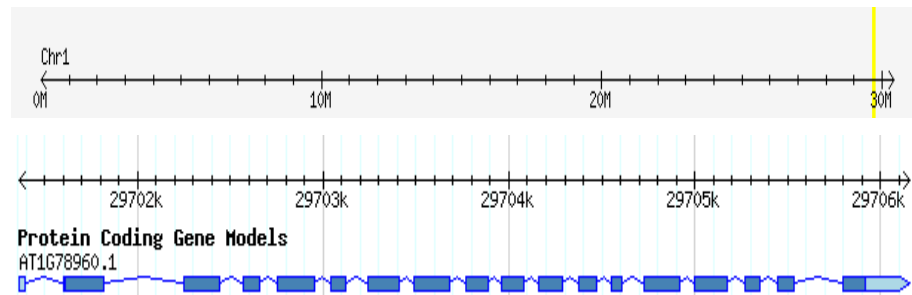
Beta-amyrin synthase



Number of exons retrieved: 17

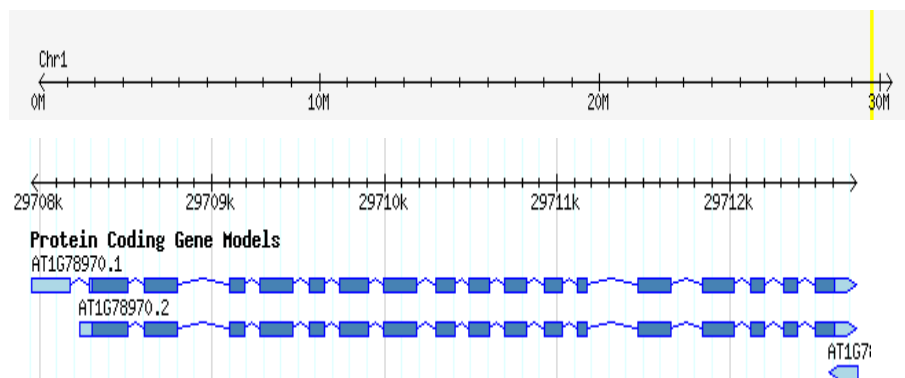
#	Exon Contents	5' End	3' End	+ strand	- strand	Size (bp)	Beginning of start
1	ORF	29,693,566	29,693,363	27,29,31	28,30,32	203	
Intron							
2	ORF	29,693,034	29,692,849			185	
Intron							
3	ORF	29,692,616	29,692,527			89	
Intron							
4	ORF	29,692,449	29,692,252	21	22	197	
Intron							
5	ORF	29,692,169	29,692,085	17	18	84	
Intron							
6	ORF	29,691,929	29,691,763			166	
Intron							
7	ORF	29,691,675	29,691,484			191	
Intron							
8	ORF	29,691,415	29,691,302			113	
Intron							
9	ORF	29,691,216	29,691,097			119	
Intron							
10	ORF	29,691,006	29,690,884	15	16	122	
Intron							
11	ORF	29,690,800	29,690,702			98	
Intron							
12	ORF	29,690,617	29,690,561			56	
Intron							
13	ORF	29,690,466	29,690,276	11,13	12,14	190	
Intron							
14	ORF	29,690,190	29,690,013			177	
Intron							
15	ORF	29,689,935	29,689,855			80	
Intron							
16	ORF	29,689,771	29,689,688	3	4	83	
Intron							
17	ORF + 3' UTR	29,689,561	29,689,325			236	
Within 3' UTR							

## Multifunctional oxidosqualene cyclase



Number	of	Exon	5' End	3' End	+ strand	-	Size
1		5' UTR	29,701,363	29,701,395			32
Intron					1	2	
2		5' UTR + ORF	29,701,601	29,701,818	3,5,7	4,6,8	217
Intron					9	10	
3		ORF	29,702,253	29,702,438	11,13	12,14	185
Intron							
4		ORF	29,702,568	29,702,657			89
Intron					15	16	
5		ORF	29,702,754	29,702,951			197
Intron							
6		ORF	29,703,039	29,703,123			84
Intron					17	18	
7		ORF	29,703,243	29,703,409			166
Intron							
8		ORF	29,703,492	29,703,683			191
Intron							
9		ORF	29,703,771	29,703,884	19,21	20,22	113
Intron							
10		ORF	29,703,961	29,704,080			119
Intron							
11		ORF	29,704,162	29,704,284	23,25	24,26	122
Intron							
12		ORF	29,704,373	29,704,471	27	28	98
Intron							
13		ORF	29,704,554	29,704,610			56
Intron					29	30	
14		ORF	29,704,725	29,704,915	31,33	32,34	190
Intron							
15		ORF	29,704,998	29,705,175			177
Intron					35	36	
16		ORF	29,705,271	29,705,351	37	38	80
Intron							
17		ORF	29,705,448	29,705,531	39	40	83
Intron					41,43	42,44	
18		ORF + 3' UTR	29,705,795	29,706,156	45,47,48	46,49	361
Within 3' UTR							

## Lupeol synthase

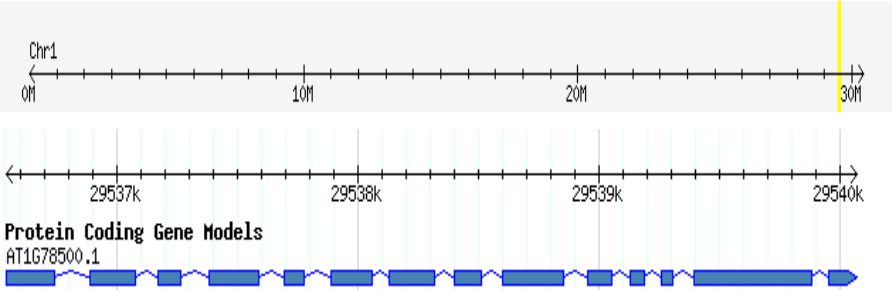


Number of exons retrieved: 18

#	Exon Contents	5' End	3' End	+ strand	- strand	Size (bp)	Beginning of start codon:
1	5' UTR	29,707,955	29,708,176	1	2	221	
Intron							
2	5' UTR + ORF	29,708,290	29,708,510	3,5,7	4,6,8	220	
Intron							
3	ORF	29,708,613	29,708,798			185	
Intron							
4	ORF	29,709,103	29,709,192			89	
Intron							
5	ORF	29,709,278	29,709,466	9	10	188	
Intron							
6	ORF	29,709,567	29,709,651			84	
Intron							
7	ORF	29,709,744	29,709,910			166	
Intron							
8	ORF	29,709,996	29,710,187			191	
Intron							
9	ORF	29,710,298	29,710,411			113	
Intron				11	12		
10	ORF	29,710,492	29,710,611			119	
Intron							
11	ORF	29,710,699	29,710,821	13,15	14,16	122	
Intron							
12	ORF	29,710,929	29,711,027			98	
Intron							
13	ORF	29,711,122	29,711,178			56	
Intron							
14	ORF	29,711,470	29,711,660	17,19,21	18,20,22	190	
Intron				23	24		
15	ORF	29,711,847	29,712,024			177	
Intron							
16	ORF	29,712,121	29,712,201	25	26	80	
Intron							
17	ORF	29,712,311	29,712,394			83	
Intron				27,29,31	28,30,32		
18	ORF + 3' UTR	29,712,495	29,712,737			242	

Within 3' UTR	33		
---------------	----	--	--

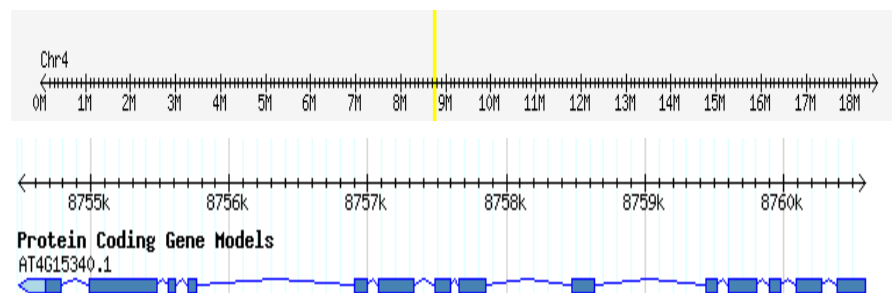
### Multifunctional including seco amyrin



Number of exons retrieved: 14

#	Exon Contents	5' End	3' End	+ strand	- strand	Size (bp)Beginning of start codon:
1	ORF	29,536,539	29,536,742	1,3,5,7	2,4,6,8	203
Intron				9	10	
2	ORF	29,536,888	29,537,073			185
Intron						
3	ORF	29,537,171	29,537,260			89
Intron				11,13	12,14	
4	ORF	29,537,383	29,537,586	15,17,19,21	16,18,20,22	203
Intron						
5	ORF	29,537,693	29,537,777			84
Intron				23	24	
6	ORF	29,537,890	29,538,056			166
Intron						
7	ORF	29,538,127	29,538,318	25	26	191
Intron						
8	ORF	29,538,398	29,538,511			113
Intron						
9	ORF	29,538,599	29,538,853	27	28	254
Intron						
10	ORF	29,538,955	29,539,053			98
Intron						
11	ORF	29,539,129	29,539,185			56
Intron						
12	ORF	29,539,257	29,539,303			46
Intron						
13	ORF	29,539,393	29,539,879	29	30	486
Intron						
14	ORF	29,539,954	29,540,070			116

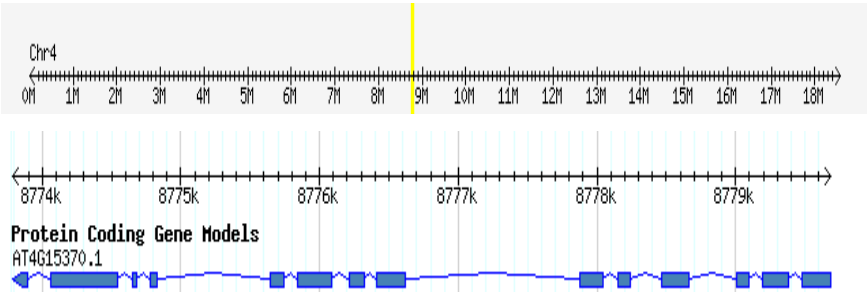
### TTPS2 arabidiol synthase



Number of exons retrieved: 14

#	Exon Contents	5' End	3' End	+ strand	- strand	Size (bp)	Beginning of start codon:
1	ORF	8,760,589	8,760,386	30	31	203	
	Intron						
2	ORF	8,760,271	8,760,086			185	
	Intron						
3	ORF	8,759,983	8,759,894			89	
	Intron						
4	ORF	8,759,804	8,759,601	26,28	27,29	203	
	Intron						
5	ORF	8,759,522	8,759,438			84	
	Intron						
6	ORF	8,758,635	8,758,469			166	
	Intron						
7	ORF	8,757,853	8,757,662	17,20	18,19,21	191	
	Intron						
8	ORF	8,757,592	8,757,479			113	
	Intron						
9	ORF	8,757,329	8,757,078	9,11,13	10,12,14	251	
	Intron						
10	ORF	8,756,997	8,756,899			98	
	Intron						
11	ORF	8,755,760	8,755,704			56	
	Intron						
12	ORF	8,755,618	8,755,572			46	
	Intron						
13	ORF	8,755,475	8,754,989	1,3	2,4	486	
	Intron						
14	ORF + 3' UTR	8,754,786	8,754,480			306	
	Within 3' UTR						

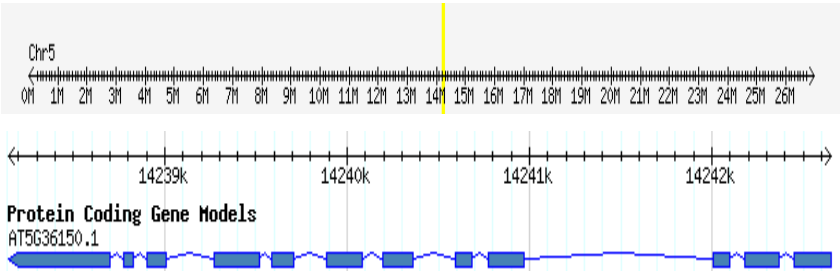
**TPPS3 Baruol synthase**



Number of exons retrieved: 14

#	Exon Contents	5' End	3' End	+ strand	- strand	Size (bp)	Beginning of start codon:
1	ORF	8,779,685	8,779,482	23	24	203	
Intron							
2	ORF	8,779,383	8,779,198			185	
Intron							
3	ORF	8,779,092	8,779,003			89	
Intron							
4	ORF	8,778,666	8,778,466	15,17,19	16,18,20	200	
Intron							
5	ORF	8,778,236	8,778,152			84	
Intron							
6	ORF	8,778,044	8,777,878			166	
Intron							
7	ORF	8,776,615	8,776,415	11,13	12,14	200	
Intron							
8	ORF	8,776,325	8,776,212			113	
Intron							
9	ORF	8,776,090	8,775,839			251	
Intron							
10	ORF	8,775,744	8,775,646			98	
Intron							
11	ORF	8,774,823	8,774,775			48	
Intron							
12	ORF	8,774,680	8,774,647			33	
Intron							
13	ORF	8,774,546	8,774,060	3,5,7	4,6,8	486	
Intron							
14	ORF	8,773,896	8,773,786			110	

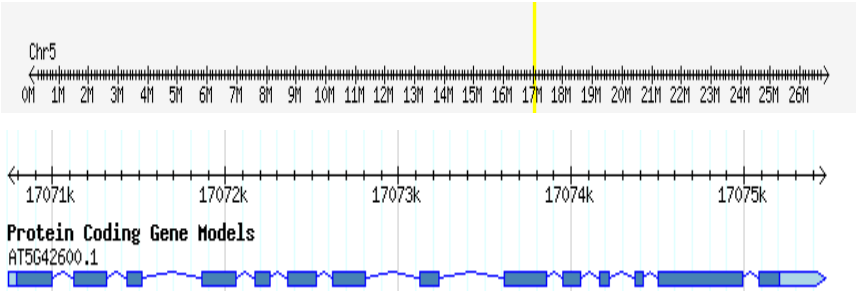
**TPPS4**



Number of exons retrieved: 12

#	Exon Contents	5' End	3' End	+ strand	- strand	Size (bp)Beginning of start codon:
1	ORF	14,242,652	14,242,449	33,35,37	34,36,38	203
Intron						
2	ORF	14,242,368	14,242,183			185
Intron						
3	ORF	14,242,096	14,242,007			89
Intron						
4	ORF	14,240,970	14,240,773	17,19,21	18,20,22	197
Intron						
5	ORF	14,240,681	14,240,597			84
Intron						
6	ORF	14,240,361	14,240,195			166
Intron						
7	ORF	14,240,082	14,239,891	13	14	191
Intron						
8	ORF	14,239,703	14,239,590	9	10	113
Intron						
9	ORF	14,239,515	14,239,273	7	8	242
Intron						
10	ORF	14,239,004	14,238,906	5	6	98
Intron						
11	ORF	14,238,830	14,238,774			56
Intron						
12	ORF	14,238,698	14,238,144	3	4	554

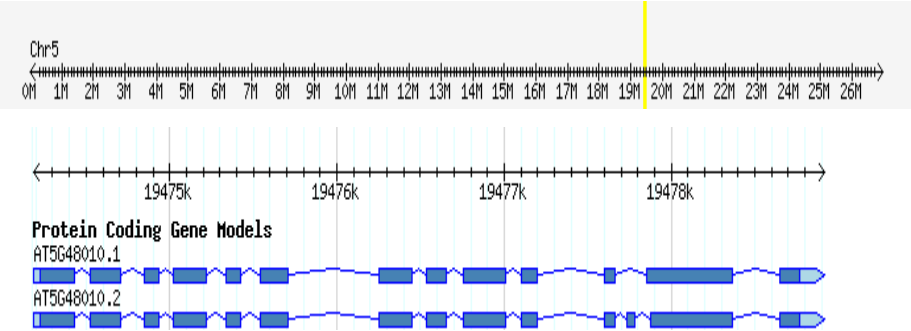
**TPPS5 Maneral synthase**



Number of exons retrieved: 14

#	Exon Contents	5' End	3' End	+ strand	- strand	Size (bp)	Beginning of start codon:
1	5' UTR + ORF	17,070,745	17,070,997	1,3	2,4	252	
Intron							
2	ORF	17,071,122	17,071,307	5	6	185	
Intron							
3	ORF	17,071,426	17,071,515			89	
Intron				7	8		
4	ORF	17,071,863	17,072,063	9,11	10,12	200	
Intron							
5	ORF	17,072,171	17,072,255			84	
Intron							
6	ORF	17,072,361	17,072,527	13	14	166	
Intron				15	16		
7	ORF	17,072,616	17,072,807	17	18	191	
Intron							
8	ORF	17,073,122	17,073,235	19	20	113	
Intron				21	22		
9	ORF	17,073,615	17,073,857	23,25	24,26	242	
Intron							
10	ORF	17,073,953	17,074,051			98	
Intron							
11	ORF	17,074,161	17,074,217			56	
Intron							
12	ORF	17,074,369	17,074,415			46	
Intron				27	28		
13	ORF	17,074,504	17,074,990	29,31	30,32	486	
Intron							
14	ORF + 3' UTR	17,075,090	17,075,473	33	34	383	
Within 3' UTR							

**TTPS6**

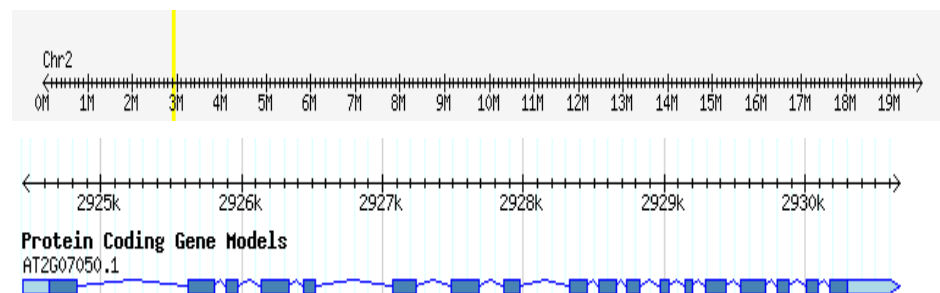




Number of exons retrieved: 13

#	Exon Contents	5' End	3' End	+ strand	- strand	Size (bp)Beginning of start codon:
1	5' UTR + ORF	19,474,186	19,474,430	1	2	244
Intron				3	4	
2	ORF	19,474,524	19,474,709			185
Intron						
3	ORF	19,474,848	19,474,937			89
Intron						
4	ORF	19,475,019	19,475,219	5,7,9	6,8,10	200
Intron				11,13,15	12,14,16	
5	ORF	19,475,338	19,475,422			84
Intron						
6	ORF	19,475,541	19,475,707			166
Intron				17	18	
7	ORF	19,476,254	19,476,445	19,21	20,22	191
Intron						
8	ORF	19,476,538	19,476,651			113
Intron						
9	ORF	19,476,757	19,477,011	23,25,27,29	24,26,28,30	254
Intron						
10	ORF	19,477,100	19,477,198			98
Intron						
11	ORF	19,477,604	19,477,660			56
Intron						
12	ORF	19,477,853	19,478,362	31,33,35	32,34,36	509
Intron						
13	ORF + 3' UTR	19,478,648	19,478,915			267
Within 3' UTR						

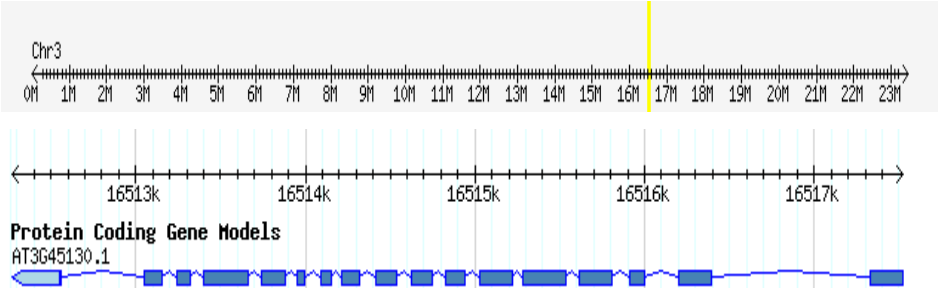
## CAS1



Number of exons retrieved: 18

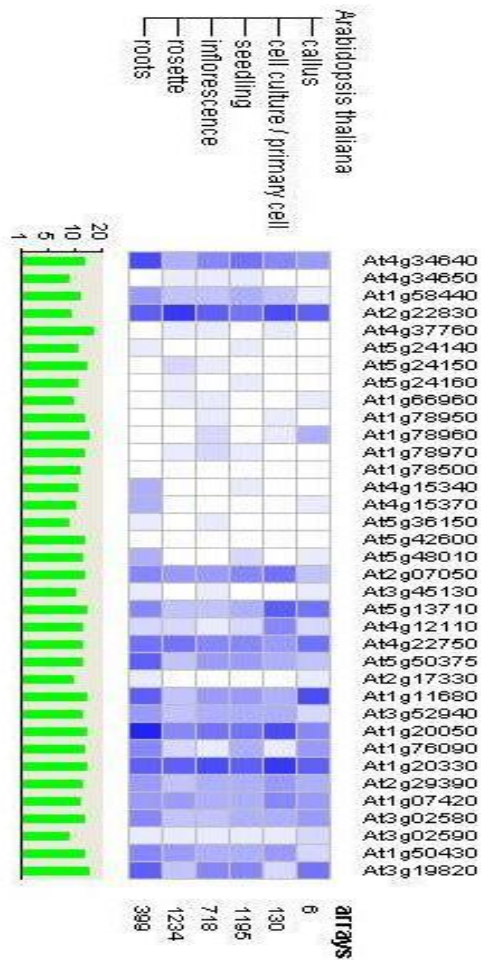
#	Exon Contents	5' End	3' End	+ strand	- strand	Size (bp)	Beginning of start codon:
1	5' UTR + ORF	2,924,444	2,924,823	1,3,5,7,9,11	2,4,6,8,10,12	379	
Intron				13,15	14,16		
2	ORF	2,925,618	2,925,803			185	
Intron							
3	ORF	2,925,883	2,925,972			89	
Intron							
4	ORF	2,926,139	2,926,333			194	
Intron				17	18		
5	ORF	2,926,437	2,926,521			84	
Intron				19,21	20,22		
6	ORF	2,927,065	2,927,231			166	
Intron				23	24		
7	ORF	2,927,488	2,927,676			188	
Intron							
8	ORF	2,927,857	2,927,970			113	
Intron							
9	ORF	2,928,324	2,928,446			122	
Intron							
10	ORF	2,928,531	2,928,653	25,27	26,28	122	
Intron							
11	ORF	2,928,728	2,928,826			98	
Intron				29	30		
12	ORF	2,928,971	2,929,027			56	
Intron							
13	ORF	2,929,146	2,929,192			46	
Intron							
14	ORF	2,929,287	2,929,430			143	
Intron							
15	ORF	2,929,540	2,929,717	31	32	177	
Intron							
16	ORF	2,929,801	2,929,881	33	34	80	
Intron							
17	ORF	2,930,007	2,930,090			83	
Intron							
18	ORF + 3' UTR	2,930,173	2,930,669	35,36	37	496	
Within 3' UTR							

LAS



Number of exons retrieved: 17

#	Exon Contents	5' End	3' End	+ strand	- strand	Size (bp)	Beginning of start codon: 16,528,507 bp
1	ORF	16,528,507	16,528,319	37	38	188	
Intron							
2	ORF	16,527,378	16,527,184			194	
Intron							
3	ORF	16,526,983	16,526,894	29	30	89	
Intron							
4	ORF	16,526,795	16,526,601			194	
Intron							
5	ORF	16,526,516	16,526,265			251	
Intron							
6	ORF	16,526,202	16,526,014			188	
Intron							
7	ORF	16,525,925	16,525,812			113	
Intron							
8	ORF	16,525,736	16,525,614	25	26	122	
Intron							
9	ORF	16,525,526	16,525,404	19	20	122	
Intron							
10	ORF	16,525,301	16,525,203			98	
Intron							
11	ORF	16,525,137	16,525,081	17	18	56	
Intron							
12	ORF	16,524,992	16,524,946			46	
Intron							
13	ORF	16,524,868	16,524,725			143	
Intron							
14	ORF	16,524,644	16,524,386			258	
Intron							
15	ORF	16,524,307	16,524,224	13,15	14,16	83	
Intron							
16	ORF	16,524,140	16,524,034	11	12	106	
Intron							
17	ORF + 3' UTR	16,523,540	16,523,256	1,4	2,3,5,6	284	
Within 3' UTR							



Tissue specific gene expression (Darkest colour top 1% of signal), of triterpene synthases displayed as a heat map. A comparison between tissue groups; callus, cell culture, seedling, inflorescence, rosette, roots. The expression patterns of genes involved in the phytosterol pathway are shown.

**APPENDIX IV: CROSS HYBRIDISATION ASSESSMENT USING BLAST SEARCHES WITHIN AFFYMETRIX ONLINE TOOL RESOURCE**

Squalene synthase (At4g34640) probe ID 253206\_at individual probe sequences 1-11 (NetAffx) and BLASTN (TAIR) sequence and score (E-value).

<b>Squalene synthase</b>	<b>AT4G34640</b>	<b>253206_at</b>
<b>Probe Sequence(5'-3')</b>	<b>AGI Code</b>	<b>E-Value</b>
GAGATACCAAAATCCCGCATGTTTT	AT4G34640	5.00E-07
	AT5G48460	0.47
TTTGGCCTCGCGAGATTGTTGGGCAA	AT4G34640	50 5e-07
	AT4G34650	0.008
ATACATGGTTTCCTTGCGTGATCCT	AT4G34640	5.00E-07
	AT4G03490	0.47
TGATCCTTCCATATTTCGGTTCTGT	AT4G34640	5.00E-07
	AT5G48990	1.9
GTGCCATCCCTCAGATCATGGCGAT	AT4G34640	5.00E-07
	AT4G34650	0.002
GAGGCGAGGTCTTACTGCTAAAGTC	AT4G34640	0.0000005
	AT5G06940	1.9
TGGTGCTTTCTATGATTTTTCCTGC	AT4G34640	0.0000005
	AT5G06940	1.9
ACTAAACCGACTTGAAGCCGTTTCAG	AT4G34640	0.0000005
	AT5G13710	1.9
TCTGCAGAGACGCTGGAGTTCTTCA	AT4G34640	0.0000005
	AT5G41100	0.47
GGTTGTGATTCTACTGGCCATAGTC	AT4G34640	0.0000005
	AT4G34650	0.002
TGGCCATAGTCTTTGCATATCTCAG	AT4G34640	0.0000005
	AT4G34650	0.12

Squalene synthase (At4g34650) probe ID 253254\_at individual probe sequences 1-11 (NetAffx) and BLASTN (TAIR) sequence and score (E-value).

<b>Squalene synthase</b>	<b>AT4G34650</b>	<b>253254_at</b>
<b>Probe Sequence(5'-3')</b>	<b>AGI Code</b>	<b>E-Value</b>
GTCAAAAATCTTCATTGCTTCGGAA	AT4G34650	0.0000005
	AT5G21222	0.12
AAATTCTACAGGTTTATTTCTGCAG	AT4G34650	0.0000005
	AT4G34640	0.47
GAGAGACCAAAGTCGCGCATGTTTT	AT4G34650	0.0000005
	AT4G22930	1.9
TTTTGGCCTCGTGAGATTTGGGGAA	AT4G34650	0.0000005
	AT4G34640	0.002
GTTTGAAATCCTTGGCTTCACTGCG	AT4G34650	0.0000005
	AT4G09730	0.47
GCTTCACTGCGTGATCCTGCAATAT	AT4G34650	0.0000005
	AT1G34844	0.47
GCAATATTTCAGTCTTGCGCCATCC	AT4G34650	0.0000005
	AT4G08262	0.47
ATGGATGATGTCTACGGTGCGTTCT	AT4G34650	0.0000005
	AT3G28750	0.47
ACGGTGCGTTCTATGATTTTTCTTG	AT4G34650	0.0000005
	AT2G38290	0.03
GATTTTTCTTGCA TGCTACAAACAA	AT4G34650	0.0000005
	AT5G24775	1.9
CATTAAACCGACTCGAAACCATCAA	AT4G34650	0.0000005
	AT1G67050	0.47

# Squalene epoxidases

Squalene epoxidase (At1g58440) probe ID 245809\_at individual probe sequences 1-11 (NetAffx) and BLASTN (TAIR) sequence and score (E-value).

<b>Squalene epoxidase</b>	<b>AT1G58440</b>	<b>245809_at</b>
<b>Probe Sequence(5'-3')</b>	<b>AGI code</b>	<b>E-value</b>
TCCTACTCCAGGGGCTCTGTTAATG	AT1G58440	5.00E-07
	AT3G51050	1.9
TGACATTGTTGTCCTGCGTAATCTC	AT1G58440	5.00E-07
	AT5G61490	1.9
TAGACCGCTGCGTGATCTTAGTGAC	AT1G58440	5.00E-07
	AT5G28710	1.9
TCAACACCCTTGCGAATGCTCTTTA	AT1G58440	5.00E-07
	AT2G34530	1.9
GGACCAGTATCTTTGCTTTCGGGTT	AT1G58440	5.00E-07
	AT1G06210	0.008
GAACCCTCGACCATTAACACTTGTC	AT1G58440	5.00E-07
	AT4G36860	1.9
TAACACTTGTCTGCCATTTCTTTGC	AT1G58440	5.00E-07
	AT5G60400	0.47
TCATACGGTTGTTAATCCCATTCCC	AT1G58440	5.00E-07
	AT5G08100	0.47
TCCCTTCCCCAAAACGAATCTGGCT	AT1G58440	5.00E-07
	AT5G12260	0.47
GTTAGGCAGATGTTTTTCCCAGCAA	AT1G58440	5.00E-07
	AT5G32433	1.9
TTCCCAGCAACTGTACCTGCATACT	AT1G58440	5.00E-07
	AT5G43960	47

Squalene epoxidase (At5g24140) probe ID 249773\_at individual probe sequences 1-11 (NetAffx) and BLASTN (TAIR) sequence and score (E-value).

<b>Squalene epoxide</b>	<b>AT5G24140</b>	<b>249773</b>
<b>Probe Sequence(5'-3')</b>	<b>AGI code</b>	<b>E value</b>
TTCATACACCTGGAGCTTTGCTTTT	AT4G37760	5.00E-07
	AT1G49810	0.47
AGGTGATGCGTTCAACATGCGCCAT	AT4G37760	5.00E-07
	AT5G45116	1.9
TGCGCCATCCTCTTACTGGAGGTGG	AT4G37760	5.00E-07
	AT5G20280	1.9
GTCATCCTCCGTGATCTATTGAACC	AT4G37760	5.00E-07
	AT1G06770	0.12
AAACCGGTTGCTTCGACTATCAATA	AT4G37760	5.00E-07
	AT4G11690	0.47
TCAATACACTCGCAGGCGCTCTTTA	AT4G37760	5.00E-07
	ATMG01090	7.3
TTGCTTCGATTATCTTAGCCTCGGA	AT4G37760	5.00E-07
	AT2G22830	0.03
CATCAGGACCTGTGGCTTTGCTATC	AT4G37760	5.00E-07
	AT5G61070	1.9
GCTATCTGGTTTGAACCCACGACCT	AT4G37760	5.00E-07
	AT5G41605	1.9
TTCTTCATTTCTTCGCAGTTGCGAT	AT4G37760	5.00E-07
	AT2G27550	0.47
TATTCCTGCCATTTACAGAGCTCCT	AT4G37760	5.00E-07
	AT4G09810	1.9



Squalene epoxidase (At2g22830) probe ID 266831\_at individual probe sequences 1-11 (NetAffx) and BLASTN (TAIR) sequence and score (E-value).

Squalene epoxidase	AT2G22830	266831_at
Probe Sequence(5'-3')	AGI code	E value
GTGGGATGACCGTTGCATTGGCGGA	AT2G22830	5.00E-07
	AT4G37760	0.12
TGTACTCCGTGATCTTCTAAGGCCA	AT2G22830	5.00E-07
	AT5G40605	1.9
AGGCCAATTCGCAACCTTAATGACA	AT2G22830	5.00E-07
	AT1G74530	0.47
AACCTGTAGCTTCCACCATTAATAC	AT2G22830	5.00E-07
	AT4G18690	0.47
CTTGCTTCGACTATCTTAGCCTTGG	AT2G22830	5.00E-07
	AT4G37760	0.008
TCATCTGGTCCAGTTGCATTGCTCT	AT2G22830	5.00E-07
	AT4G04970	0.47
GCATTGCTCTCTGGTTTAAACCCTC	AT2G22830	5.00E-07
	AT5G41100	0.12
TCGTCCTCTGAGTTTAGTTCTCCAC	AT2G22830	5.00E-07
	AT5G16120	0.47
TTCTCCACTTCTTTGCTGTGGCGAT	AT2G22830	5.00E-07
	AT3G59320	0.47
GCGATCTACGCTGTTTGTTCGTTTAA	AT2G22830	5.00E-07
	AT4G33470	1.9
GAGTGCTTCAAGCATCATCTTTCCA	AT2G22830	5.00E-07
	AT4G11720	0.47

Squalene epoxidase (At5g24150) probe ID 249774\_at individual probe sequences 1-11 (NetAffx) and BLASTN (TAIR) sequence and score (E-value).

<b>Squalene epoxidase</b>	<b>AT5G24150</b>	<b>249774_at</b>
<b>Probe Sequence(5'-3')</b>	<b>AGI code</b>	<b>E value</b>
TCAACATGCGTCATCCAGCAATCGC	AT5G24150	5.00E-07
	AT3G17160	0.47
ATGCATTCTCTCAAGTGCTAGTTGC	AT5G24150	5.00E-07
	AT4G02940	0.47
GTTGCTATGATTACCTCTCTAGTGG	AT5G24150	5.00E-07
	AT5G45200	1.9
TCTAGTGGTGGGTTTCGCACGTCAG	AT5G24150	5.00E-07
	AT2G25740	0.47
AGGGATGATGGCTTTGCTAGGCGGC	AT5G24150	5.00E-07
	AT5G24160	0.03
ATCTCTCTCATCTATCATCTATGTG	AT5G24150	5.00E-07
	AT4G06735	0.47
ATCTATGTGCTATCACTCTATCCTC	AT5G24150	5.00E-07
	AT5G24160	0.03
CATTTGGCATAGCCTTCGACTTTTT	AT5G24150	5.00E-07
	AT5G24160	0.47
GTTCCCATCTCAAGGCTGAAGGAG	AT5G24150	5.00E-07
	AT5G24160	0.03
AAATGTTGTTTCCAGTCAACGCCGC	AT5G24150	5.00E-07
	AT4G23350	0.12
GTCAACGCCGCCGCGTATAGCAAAA	AT5G24150	5.00E-07
	AT5G44550	7.3

Squalene epoxidase (At4g37760) probe ID 253039\_at individual probe sequences 1-11 (NetAffx) and BLASTN (TAIR) sequence and score (E-value).

<b>Squalene epoxidase</b>	<b>AT4G37760</b>	<b>253039_at</b>
<b>Probe Sequence(5'-3')</b>	<b>AGI code</b>	<b>E value</b>
TTCATACACCTGGAGCTTTGCTTTT	AT4G37760	5.00E-07
	AT1G49810	0.47
AGGTGATGCGTTCAACATGCGCCAT	AT4G37760	5.00E-07
	AT5G45116	1.9
TGCGCCATCCTCTTACTGGAGGTGG	AT4G37760	5.00E-07
	AT5G20280	1.9
GTCATCCTCCGTGATCTATTGAACC	AT4G37760	5.00E-07
	AT1G06770	0.12
AAACCGGTTGCTTCGACTATCAATA	AT4G37760	5.00E-07
	AT4G11690	0.47
TCAATACACTCGCAGGCGCTCTTTA	AT4G37760	5.00E-07
	ATMG01090	7.3
TTGCTTCGATTATCTTAGCCTCGGA	AT4G37760	5.00E-07
	AT2G22830	0.03
CATCAGGACCTGTGGCTTTGCTATC	AT4G37760	5.00E-07
	AT5G61070	1.9
GCTATCTGGTTTGAACCCACGACCT	AT4G37760	5.00E-07
	AT5G41605	1.9
TTCTTCATTTCTTCGCAGTTGCGAT	AT4G37760	5.00E-07
	AT2G27550	0.47
TATTCCTGCCATTTACAGAGCTCCT	AT4G37760	5.00E-07
	AT4G09810	1.9

Squalene epoxidase (At5g24160) probe ID 249775\_at individual probe sequences 1-11 (NetAffx) and BLASTN (TAIR) sequence and score (E-value).

<b>Squalene epoxidase</b>	<b>AT5G24160</b>	<b>249775_at</b>
<b>Probe Sequence(5'-3')</b>	<b>AGI code</b>	<b>E value</b>
ATGCGTCATCCAGTTGTTGCATCTG	AT5G24160	5.00E-07
	AT5G10310	0.47
GGAATGATGGTTTTACTGTCTGGACA	AT5G24160	5.00E-07
	AT5G24150	0.47
GCCATTAAGCAACCTCGGCGATGCA	AT5G24160	5.00E-07
	AT5G24150	0.12
CCGCAAGCCAATGTCTGGCGACGGTT	AT5G24160	5.00E-07
	AT5G24150	0.03
GTCGGCGACGGTTAACACATTGGGA	AT5G24160	5.00E-07
	AT5G43030	1.9
GGGTGTCTATGATTACCTTTGTAGT	AT5G24160	5.00E-07
	AT3G07160	1.9
TTGTAGTGGCGGGTTTCGTACGTCA	AT5G24160	5.00E-07
	AT5G15820	7.3
GGCATAGCCTCAAGCTTTTTGGTTT	AT5G24160	5.00E-07
	AT5G24310	0.12
TGTTGGTTCCCAATCTCAAAGCTGA	AT5G24160	5.00E-07
	AT5G24150	0.03
GTTGTTTCCAGCAAATGCAGCCGCG	AT5G24160	5.00E-07
	AT4G34220	0.47
TGCAGCCGCGTATCACAAAAGCTAT	AT5G24160	5.00E-07
	AT2G14260	0.47

# **Oxidosqualene cyclases**

Oxidosqualene cyclase (At1g66960) probe ID 2255912\_at individual probe sequences 1-11 (NetAffx) and BLASTN (TAIR) sequence and score (E-value).

<b>Lupeol synthase</b>	<b>AT1G66960</b>	<b>255912_AT</b>
<b>Probe Sequence(5'-3')</b>	<b>AGI code</b>	<b>E value</b>
AGCTGGGGTATTTGTTTCACTTATG	AT1G66960	7.00E-07
	AT5G03340	0.043
TTATGGGACATGGTTTGCTCTTTGC	AT1G66960	7.00E-07
	AT3G13750	0.17
GTCTATCTATGCGCGACGGTGTACA	AT1G66960	7.00E-07
ACGGTGTACATTTCTTCTTAATAT	AT1G66960	7.00E-07
	AT4G27010	0.17
GAAAGCTATATGTCATGCCCTGAAC	AT1G66960	7.00E-07
	AT3G06810	0.17
AAACGTAGTGCAAACCGCGTGGGCT	AT1G66960	7.00E-07
	AT1G78960	<u>0.043</u>
GGCTATGATGGCTCTGATTCACGCT	AT1G66960	7.00E-07
	AT1G78979	<u>0.043</u>
AGATCTTATACCTCTACATAGTGCT	AT1G66960	7.00E-07
	AT4G06508	2.7
CATACAAAGACATCTTCCCACCATG	AT1G66960	7.00E-07
	AT5G18245	0.17
GGGCACTTGCAGAGTACCGGAAAGC	AT1G66960	7.00E-07
	AT1G16600	0.67
CCGGAAAGCTGCGTTCATACATCAC	AT1G66960	7.00E-07
	AT4G06654	0.17

Oxidosqualene cyclase (At1g78950) probe ID 264138\_at individual probe sequences 1-11 (NetAffx) and BLASTN (TAIR) sequence and score (E-value).

<b>Beta-amyrin synthase</b>	<b>AT1G78950</b>	<b>264138_at</b>
<b>Probe Sequence(5'-3')</b>	<b>AGI code</b>	<b>E value</b>
GAATGTACATCGTCTGCAATCCAAG	AT1G78950	7.00E-07
	AT5G66120	2.7
TTTTCAAGCAACTCTACCCTGATCA	AT1G78950	7.00E-07
	AT3G03360	0.17
GGACAACAGAGATCACCGCTTTCAT	AT1G78950	7.00E-07
	AT3G60400	0.67
GAACTGGGGCATTGCTTCACGTAC	AT1G78950	7.00E-07
	AT1G26170	0.67
TACGGTACATGGTTTGCTCTTGCA	AT1G78950	7.00E-07
	AT1G66960	0.67
GCTCTTGCAAGCTTAGCAGCTGCGG	AT1G78950	7.00E-07
	AT4G37460	0.67
TGGGGAGAAAGCTACCTCTCTTGCT	AT1G78950	5.00E-07
	AT5G58250	0.47
GTGGTGCAAAGCTTGGGCTTTAA	AT1G78950	5.00E-07
	AT5G04930	0.47
GATTTTCCTCAACAGCAAGCAACCG	AT1G78950	5.00E-07
	AT5G45150	0.47
GAAACATTCATCCGTTGTGGGCACT	AT1G78950	5.00E-07
	AT4G39753	1.9
AATATCGCGCGAGTTTCGTTGCC	AT1G78950	5.00E-07
	AT4G00560	1.9

Oxidosqualene cyclase (At1g78960) probe ID 264137\_at individual probe sequences 1-11 (NetAffx) and BLASTN (TAIR) sequence and score (E-value).

<b>Lupeol synthase</b>	<b>AT1G78960</b>	<b>264137_at</b>
Probe Sequence(5'-3')	AGI code	E value
GTTCAAACAAC TTTATCCGGATCAC	AT1G78960	5.00E-07
	AT1G66960	0.47
ACTGGGGTATCTGTTTCATCTACGC	AT1G78960	5.00E-07
	AT5G32433	0.47
TACTTG GTTTGCTCTGAGCGGCCTA	AT1G78960	5.00E-07
	AT2G41312	0.002
AAAAGTTGCCTGGCTGTGCGCAAAG	AT1G78960	5.00E-07
	AT5G46665	0.47
AAGGTGTAGATTTCTGCTTGCGAT	AT1G78960	5.00E-07
	AT5G21150	1.9
GGGCGAAAGCCATCTGTCATGCCCT	AT1G78960	5.00E-07
	AT1G17680	0.47
AACCTAGTGCAAACCGCATGGGCTA	AT1G78960	5.00E-07
	AT1G66960	0.03
GATTCATGCCGGACAGGCCGAGAGA	AT1G78960	5.00E-07
	AT1G78955	0.12
CAGGCCGAGAGAGATCCTACACCTC	AT1G78960	5.00E-07
	AT2G27040	0.008
GAATATCGGAAAGCTGCCTTCGCAA	AT1G78960	5.00E-07
	AT5G28800	0.12
GCCTTCGCAACTCATCAAGATCTTT	AT1G78960	5.00E-07
	AT1G02750	0.47

Oxidosqualene cyclase (At1g78970) probe ID 264100\_at individual probe sequences 1-11 (NetAffx) and BLASTN (TAIR) sequence and score (E-value).

<b>Lupeol synthase</b>	<b>AT1G78970</b>	<b>264100_at</b>
Probe Sequence(5'-3')	AGI code	E value
GTGGAATGCACCTCATCTGTTATAC	AT1G78970	5.00E-07
	AT3G58810	1.9
GGTGTTTGCTTCATTTACGCTACTT	AT1G78970	5.00E-07
	AT5G39940	0.47
TACTTGGTTTGCTCTTGGAGGCCTA	AT1G78970	5.00E-07
	AT3G17650	0.03
TCTTGAGGCCTAGCAGCAGCTGGT	AT1G78970	5.00E-07
	AT4G08490	1.9
TAGCTATGCGCAATGGTGTCCACTT	AT1G78970	5.00E-07
	AT5G30276	1.9
GTGTCCACTTTTTGCTCACGACACA	AT1G78970	5.00E-07
	AT1G00180	1.9
GCTATTTATCATGCTCCGAACAGAG	AT1G78970	5.00E-07
	AT5G67220	1.9
AACCTTGTGCAAACATCATGGGCTA	AT1G78970	5.00E-07
	AT2G02026	0.47
GGCTCTAATTCATACGGGACAGGCT	AT1G78970	5.00E-07
	AT5G41260	1.9
GCTACATACAGAAACACCTTCCCAT	AT1G78970	5.00E-07
	AT5G40820	0.12
ACACCTTCCCATTATGGGCACTCGC	AT1G78970	5.00E-07
	AT1G78960	0.03



# ATPEN Oxidosqualene cyclases

Oxidosqualene cyclase (At1g78500) probe ID 263123\_at individual probe sequences 1-11 (NetAffx) and BLASTN (TAIR) sequence and score (E-value).

<b>Beta amyrin synthase ATPEN1</b>	<b><u>At1g78500</u></b>	<b><u>263123_at</u></b>
<b>Probe Sequence(5'-3')</b>	<b>AGI code</b>	<b>E value</b>
GCGATAGTGGCATTGGCTCGGTTCT	AT1G78950	5.00E-07
	AT5G66120	1.9
AAAGCTTTCAAATGCCCGATGGTTC	AT5G66120	5.00E-07
	AT1G78500	1.9
TGTATGGAACCTTTTTTCGCGGTAAG	AT1G07680	5.00E-07
	AT1G78500	1.9
TCGCGGTAAGGGGTCTAGTGGCTGC	AT1G78500	5.00E-07
	AT5G64120	1.9
GGCTGCAGGCAAGACGTACCAGAAC	AT1G78500	5.00E-07
	AT2G23870	0.48
GTACCAGAACTGTGAGCCGATTCGT	AT1G78500	5.00E-07
	AT5G55460	1.9
GCGGTTTCAGTTCATTCTGGAGACAC	AT1G78500	5.00E-07
	AT1G35614	0.48
GGTGAGAGTTATCTCTCTTGCCCCA	AT1G78500	5.00E-07
	AT4G15340	0.48
GCCTGTTTCATCGCGCAGCTAAAGTG	AT1G78500	5.00E-07
	AT5G42600	0.12
GTGATGGTCCATTATGCAACCTATA	AT1G78500	5.00E-07
	AT2G14500	0.48
CTACACAAAGGCGCTGCGCGTGCCC	AT1G78500	5.00E-07
	AT5G16520	7.5

Oxidosqualene cyclase (At4g15340) probe ID 245258\_at individual probe sequences 1-11 (NetAffx) and BLASTN (TAIR) sequence and score (E-value).

<b>ATPEN2 LUS - like</b>	<b>At4g15340</b>	<b>245258_at</b>
Probe Sequence(5'-3')	AGI code	E value
CAGCCATTGTAGCGTTGACTCAGTT	AT4G15340	5.00E-07
	AT5G55110	0.48
GGAGTGTGCTTTATCTATGGGACCT	AT4G15340	5.00E-07
	AT4G15370	0.031
ATCTATGGGACCTTATTTGCCGTAA	AT4G15340	5.00E-07
	AT5G05630	0.48
GACTTTCCATAACTGTGAACCCATT	AT4G15340	5.00E-07
	AT3G32112	0.12
AACCCATTCGTCGAGCAGTTCGTTT	AT4G15340	5.00E-07
	AT4G15370	0.12
AGAGCTATCTCTCTTGCCTAAGGAA	AT4G15340	5.00E-07
	AT5G06680	0.48
GTGAGTACAGGACAAGCGCTTATGG	AT4G15340	5.00E-07
	AT4G15370	0.031
GATAACGGTGATTTCCCGCAACAGG	AT4G15340	5.00E-07
	AT5G35725	0.48
TGCTGCTCCATTATCCAACATATAG	AT4G15340	5.00E-07
	AT4G15340	0.031
GGAACATTTATTCTCTTTGGGCCCT	AT5G41960	5.00E-07
	AT4G15340	0.48
TATACACGCAAGCTCTACGGCGGCT	AT4G15340	5.00E-07
	AT4G01210	1.9

Oxidosqualene cyclase (At4g15370) probe ID 245553\_at individual probe sequences 1-11 (NetAffx) and BLASTN (TAIR) sequence and score (E-value).

<b>ATPEN3 LUS like</b>	<b>At4g15370</b>	<b>245553_at</b>
Probe Sequence(5'-3')	AGI code	E value
GGATCTGCAATTGTAGCGTTGGCTC	AT4G15370	5.00E-07
	AT1G01660	1.9
AATTGGGGAGTGTGCTTCATCTATG	AT4G15370.	5.00E-07
	AT4G15340	0.008
TCATCTATGGTACCTTCTTTGCTGT	AT4G15370	5.00E-07
	AT5G55540	0.12
GTCTTGTGGCTGCTGGGAAGTGTTA	AT4G15370	5.00E-07
	AT3G20020	0.48
ATTCGTAGAGCAGTTCGTTTCATTC	AT4G15370	5.00E-07
	AT4G15340	0.031
TTCGTTTCATTCTCGACACTCAAAA	AT4G15370	5.00E-07
	AT5G17920	0.48
GAGAGCTATCTATCTTGTCCAAGAA	AT4G15370	5.00E-07
	AT1G45215	0.12
AATGGATAACGGCGATTTCCACAA	AT4G15370	5.00E-07
	AT5G36150	0.12
TGCTCCATTTTCCAACCTACAGAAA	AT4G15370	5.00E-07
	AT3G30400	0.002
CTTTGGGCTCTTACACATTACACGA	AT4G15370	5.00E-07
	AT5G48010	0.002
TTACACGAAGGCTCTGCGAGGGCTC	AT4G15370	5.00E-07
	AT2G48060	1.9

Oxidosqualene cyclase (At5g36150) probe ID 249687\_at individual probe sequences 1-11 (NetAffx) and BLASTN (TAIR) sequence and score (E-value).

<b>ATPEN4</b>	<b>At5g36150</b>	<b>249687_at</b>
Probe Sequence(5'-3')	AGI code	E value
TGAAGACACTATCCTCGAGCATGAG	AT5G42600	5.00E-07
	ATCG00160	7.5
GATAGTGGTGTGTTGGCACGCTTCATG	AT5G36150	5.00E-07
	AT5G55470	1.9
GAAACAGTTTCCAGGGCATAGAACA	AT5G36150	5.00E-07
	AT3G29255	0.031
TTGCTGTCCGAGGTTTAGTGGCCGC	AT5G36150	5.00E-07
	AT3G29255	0.48
TGGCCGCGGGAAACACTTACGATAA	AT5G36150	5.00E-07
	AT1G08300	1.9
GGCAATCCGTAGAGCAGTTCGATTC	AT5G36150	5.00E-07
	AT4G15370	0.031
GCAGTTCGATTCCTTCTTGATATAC	AT5G36150	5.00E-07
	AT3G61600	0.12
TCTTTCTTGCCCCAACAAAAATTAT	AT5G36150	5.00E-07
	AT3G08020	0.12
GGTTCACCGCGCTGCAAAAGTATTA	AT5G36150	5.00E-07
	AT4G38900	1.9
GGATAACGGTGATTTTCCACAGCAG	AT5G36150	5.00E-07
	AT5G48010	0.031
TTTCCACAGCAGGTGTGTTTCTTGT	AT5G56150	0.48
	AT5G03150	1.9

Oxidosqualene cyclase (At5g42600) probe ID 249205\_at individual probe sequences 1-11 (NetAffx) and BLASTN (TAIR) sequence and score (E-value).

<b>ATPEN5</b>	<b>At5g42600</b>	<b>249205_at</b>
Probe Sequence(5'-3')	AGI code	E value
TACCTTGAGTGTACGGGCTCGGTGA	AT5G42600	5.00E-07
	ATCG00160	7.5
AGAAAGAGTTTCCCGATCACAGACC	AT5G42600	5.00E-07
	AT5G16505	0.48
GGACTTGCAAATGCCGGACGGTTCA	AT5G42600	5.00E-07
	AT5G29040	1.9
GTTTCACGTATGGTACTCTCTTTGC	AT5G42600	5.00E-07
	AT4G31710	1.9
GTAAGAGGTCTAGCGGCTGCAGGGA	AT5G42600	5.00E-07
	AT4G15340	0.48
TGGGGAGAAAGTGCTCTCTTGCC	AT5G42600	5.00E-07
	AT5G66280	1.9
AGATGGAGAGGGACCCTTCTCCTGT	AT5G42600	5.00E-07
	AT5G11530	7.5
TTCATCGCGCCGCCAAAGTGTTGAT	AT5G42600	5.00E-07
	AT4G15370	0.031
TATGCTGCTGCATTATCCAACCTAT	AT5G42600	5.00E-07
	AT3G29255	0.002
TTGGGCTCTCGCATTGTACACAAAT	AT5G42600	5.00E-07
	AT3G62390	1.9
TGTACACAAATGCTCTGCGTCTGCT	AT5G42600	5.00E-07
	AT5G07150	0.12

Oxidosqualene cyclase (At5g48010) probe ID 248729\_at individual probe sequences 1-11 (NetAffx) and BLASTN (TAIR) sequence and score (E-value).

<b>ATPEN6 cycloartenol synthase</b>	<b>At5g48010</b>	<b>248729_at</b>
Probe Sequence(5'-3')	AGI code	E value
GCGATTGCAGCATTGACTCAGTTTA	AT5G48010	5.00E-07
	AT1G79245	0.12
GGAGTGTGTTTTATATACGGGACCT	AT5G48010	5.00E-07
	AT3G29255	0.031
TATACGGGACCTTCTTTGCGGTAAG	AT5G48010	5.00E-07
	AT3G29255	0.48
CTTGTGGCCGCTGGGAAGACTTACA	AT5G48010	5.00E-07
	AT2G37420	0.48
GAAGCAATTCGTAAAGCAGTTCGTT	AT5G48010	5.00E-07
	AT4G15370	0.008
GCAGTTCGTTTTCTTCTAGACACAC	AT5G48010	5.00E-07
	AT4G16180	0.48
GACACACAAAATCCGGAGGGTGGCT	AT5G48010	5.00E-07
	AT5G40405	0.12
GCAAACAGCACAAGCACTTATGGTG	AT5G48010	5.00E-07
	AT4G15370	0.12
ATGCTCCATTTTCCGACCTATAGGA	AT5G48010	5.00E-07
	AT4G15370	0.12
ATAGGAACACGTTCTCTCTTTGGGC	AT5G48010	5.00E-07
	AT5G42600	0.031
CACACATTACACACATGCTCTGCGA	AT5G48010	5.00E-07
	AT5G20900	1.9

Oxidosqualene cyclase (At2g07050) probe ID 266495\_at individual probe sequences 1-11 (NetAffx) and BLASTN (TAIR) sequence and score (E-value).

<b>Cycloartenol synthase</b>	<b>At2g07050</b>	<b>266495_at</b>
Probe Sequence(5'-3')	AGI code	E value
AATGTACATCAGCTGCTATCCAAGC	AT2G07050	5.00E-07
	AT5G27060	1.9
AAAGCTGTATCCTGGTCATCGAAAG	AT2G07050	5.00E-07
	AT5G43530	0.48
ATCATGGGCTGTTTGCTTCACGTAT	AT2G07050	5.00E-07
	AT4G25800	
ACTCTCCACATGTTGCTAAAGCTTG	AT2G07050	5.00E-07
	AT1G58561	0.48
TCGAAACAACAACCTTCGGGCGGCT	AT2G07050	5.00E-07
	AT2G16150	0.12
TACAGCATGGGCTATGCTCGCACTC	AT2G07050	5.00E-07
	AT3G54890	0.48
TGCTCGCACTCATTGGTGCTGGGCA	AT2G07050	5.00E-07
	AT5G49160	1.9
GCTGGGCAAGCTGAGGTAGACCGGA	AT2G07050	5.00E-07
	AT4G20450	0.48
CCACTACACCGGGCTGCAAGATACT	AT2G07050	5.00E-07
	AT3G20880	0.48
ATAACATATGCCGCGTATCGAAACA	AT2G07050	5.00E-07
	AT2G45010	1.9
GGCTTTGGGGGAGTACCGTTGTCAG	AT2G07050	5.00E-07
	AT3G43175	1.9

## APPENDIX V: AMINO ACID SEQUENCE ALIGNMENT OF SQUALENE EPOXIDASES

SQE2	TCTGGCACTGATGTGATCATCGTCGGAGCTGGTGTGCGCGGTCCGCTCTTGCTCATACT 420
SQE3	ATT-----ATCATTGTCGGTGCTGGTGTGCGCGCGCTGCCCTTGCTCATACT 274
SQE1	GTC-----ATTGTTGTGGAGCTGGTGTGCTGGTCTGCTCTTGCTTATACT 94
SQE2	CTCGGCAAGGAAGGAAGAAGAGTGCACGTTATAGAAAGAGATTTTCTGAGCAAGACAGA 480
SQE3	CTCGGCAAGGAAGGAAGAAGAGTTCACGTTATAGAAAGAGACTTAACGGAGCCTGATCGA 334
SQE1	CTTGGAAGGATAAACGCCGAGTTTCATGTGATTGAAAGAGATTTATCGGAGCCTGATCGT 154
SQE2	ATCGTTGGTGAATTGCTTCAACCTGGTGGTTATTTGAAGTTAATAGAACTTGGACTTGAA 540
SQE3	ATTGTGCGTGAATTACTTCAGCCTGGTGGTTACTTGAAGTTAATCGAACTCGGGCTTGAA 394
SQE1	ATTGTTGGGGAGTTGTTACAGCCTGGGGTTACCTCAAGTTACTGGAGTTGGGAATTGAA 214
SQE2	GATTGTGTGAAGAAGATTGATGCTCAACGAGTTCTTGGTTATGTTCTCTTTAAAGATGGG 600
SQE3	GATTGTGTGAAGGATATAGATGCGCAGAGAGTTCTTGGTTATGCTCTCTTTAAAGATGGG 454
SQE1	GATTGTGTGGAAGAAATAGATGCTCAGCGTGTGTATGGTTATGCACTTTTAAAAATGGG 274
SQE2	AAGCATACTAAACTTGCTTACCCCTTGGAAACGTTTGATTCCGATGTAGCCGGGAGAAGT 660
SQE3	AAACACACTAAACTCTCTTACCCGTTGGATCAGTTTGATTCCGATGTGCGGGTCGTAGC 514
SQE1	AAACGCATTTCGCTTAGCTTATCCTTTGGAGAAGTTTCACGAAGATGTATCTGGAAGGAGC 334
SQE2	TTCCATAATGGGAGATTGTACAGAGAATGCGAGAAAAAGCCCTTACTCTTCAATGTA 720
SQE3	TTTCACAATGGGAGATTGTGTCAGAGGATGCGAGAGAAAGCTTCTTACTTCCCAATGTT 574
SQE1	TTTACAATGGACGTTTATTCAAGAATGCGGGAGAAAGCTGCTTCACTTCCCAATGTT 394
SQE2	CGATTGGAACAAGGAACGGTTACGTCGTTGCTTGAAGAACACGGGACAATTAAGGGGTT 780
SQE3	CGAATGGAGCAAGGAACAGTGACATCGTTGGTGGAAGAAAACGGAATAATCAAAGGTGTT 634
SQE1	CAGCTAGAGCAAGGAACAGTTCTTCTCTTCTAGAAGAGAATGGGACTATCAAAGGTGTG 454
SQE2	CGATACAGAACAAAAGAGGGCAATGAGTTTAGATCATTTGCTCCTCTACAATTGTATGT 840
SQE3	CAATACAAAACCAAAGATGGCCAAGAGCTTAAGTCATTTGCTCCTCTCACTATTGTATGT 694
SQE1	AGATATAAGAATAAAGCAGGAGAGGAACAAACCGCATTTGCAGCTTTGACTATAGTTGT 514
SQE2	GATGGTTGTTTCTCCAACCTTGCCTGCTCTCTTTGCAAACCTAAGTGGAATGTGCCATCT 900
SQE3	GATGGTTGTTTCTCCAACCTTGCCTGCTCTCTTGCACCACTAAGTGGAAGGTGCCATCC 754
SQE1	GATGGTTGTTTCTCAACCTGCTGCTCTTGTGCAATCTCAGGTGCGAGGTGCTTCT 574
SQE2	ACTTTTGTGGGTCTTGTCTTGGGAACTGTGAACCTTCCATTGCAAATCACGGGCACGTT 960
SQE3	AATTCGTTGGTCTGGTATTGGAGAATTGCGAACTCCCGTTTCCGAACCAACGGGCACGTT 814
SQE1	TGTTTTGTGCGGTGGTCTAGAGAAGTGAATCTCCCATATGCAAACCATGGACATGTC 634
SQE2	GTTCTCGGTGACCCATCACCCATCTTAATGTATCCCATCAGCAGTTCTGAAGTCCGTTGC 1020
SQE3	GTTCTTGGCGATCCGTCACCCATTTTATTCTATCCTATCAGCAGCTCGGAAGTCCGTTGC 874
SQE1	GTCTTAGCAGATCCATCACCAATTTTGATGTATCCAATTAGTAGCAGAGGTGCGGTGC 694
SQE2	TTAGTAGATGTACCGGGTCAAAAACCTTCTCCCATTTGCAAATGGTGAAATGGCAAAGTAT 1080
SQE3	TTGGTAGATGTACCAGGTCAAAAACCTTCTCCCATTTGCAAAGTGGTGAAATGGCTCACCAT 934
SQE1	CTAGTTGATGTTCCCGGCCAGAAAGTCCGTCATTGCAAATGGTGAAATGAAAACTAT 754
SQE2	CTGAAAACACGGGTTGCGCCTCAAGTACCAACCAAGGTCCGTGAAGCATTCATCACCGCT 1140
SQE3	CTCAAAACAATGGTTGCACCGCAGGTACCACTCAGATCCGTGATGCTTTCATTCTGCA 994
SQE1	TTGAAGACTGTTGTGGCTCCTCAGATGCCTCATGAGGTCTATGACTCTTTCATTGCTGCG 814
SQE2	GTTGAGAAAGGTAATATCAGAACCATGCCAAACCGAAGCATGCCAGCTGATCCGATTCTCT 1200
SQE3	GTCGAAAAAGGTAACATAAGAACAATGCCGAACAGAAGCATGCCAGCTGATCCAATTAT 1054
SQE1	GTTGATAAAGGAAATATTAAGTCCATGCCAAACAGAAGCATGCCAGCTTCTCTTATCTCT 874
SQE2	ACTCCTGGAGCTCTTCTTCTTGGTGTATGCATTCAACATGAGACATCCTTTAACCGTGGT 1260
SQE3	ACACCTGGAGCTTTGCTTTTAAAGTGTATGCGTTCAACATGCGCCATCCTTCTACTGGAGGT 1114
SQE1	ACTCCAGGGGCTCTGTTAATGGGAGATGCATTTAACATGCGTCATCCTTTGACGGGTGGA 934
SQE2	GGGATGACCGTTGCATTGGCGGATATAGTTGTAATCCGTGATCTTCTAAGGCCAATTGCG 1320
SQE3	GGTATGACCGTTGCATTGTCTGATATAGTCATCCTCCGTGATCTATTGAACCCGCTCGTC 1174
SQE1	GGAATGACCGTTGCATTAGCTGACATTGTTGTCTCGTAATCTCCTTAGACCGCTGCGT 994
SQE2	AACCTTAATGACAAG--AAGCTTTGTCTAAGTATATTGAATCCTTTTACACACTACGAA 1378
SQE3	GACTTAACCAACAAAAG--AGTCTTATCCAAATACATAGAATCATTCTACACATTGCGGA 1232
SQE1	GATCTTAGTGACGGCGTAGTCTCTG--CAAATATCTTGAATCATTTTACACTCTGCGAA 1052
SQE2	AACCTGTAGCTTCCACCATTAATACATTGGCGGATGCGTTGTATAAGGTCTTTTTAGCAT 1438
SQE3	AACCGGTTGCTTCGACTATCAATACACTGCGAGGCGCTCTTTACAAAGTCTTTTTAGCAT 1292
SQE1	AGCCAGTTGCAGCAACAATCAACACCTTGCGAATGCTCTTTACCAAGTGTTCTGTTTCAT 1112
SQE2	CTTCAGATGAAGCAAGAACGGAATGCGTGAAAGCTTGCTTCGACTATCTTAGCCTTGGAG 1498
SQE3	CTCCAGACGATGCAAGAAGCGAAATGCGTCGAGCTTGCTTCGATTATCTTAGCCTCGGAG 1352
SQE1	CAGAAAATGAAGCAAGAAACGAGATGAGGGAAGCTTGCTTCGATTATCTGGGCCTCGGGG 1172
SQE2	GTGTTTTCTCATCTGGTCCAGTTGCATTGCTCTGTTTAAACCTCGTCTCTGAGTT 1558
SQE3	GGGTTTGCTCATCAGGACCTGTGGCTTTGCTATCTGGTTTGAACCCACGACCTATGAGCC 1412
SQE1	GTATGTGCACAAGTGACAGATATCTTTGCTTTCGGGTTTGAACCTCGACCATTAACAC 1232
SQE2	TAGTTCTCCACTTCTTTGCTGTGGCGATCTACGCTGTTTGTGCTTTAATGCTACCATTTC 1618



SQE3 TTGTTCTTCATTTCTTCGCAGTTGCGATTTTCGGGGTTGGTCGTTTGCTTGTACCTCTCC 1472  
 SQE1 **TGTC**TGCCATTTCTTTGCGGTTGCGGTTTATGGAGTCATACGGTTGTTAATCCCATTC 1292

SQE2 CTTTCGATTGAGAGCTTTTGGCTTGGAGCTAGGATAATCTCGAGTGCTTCAAGCATCATCT 1678  
 SQE3 CGTCCGTTAAACGGTTATGGCTTGGAGCTAGACTAATCTCGAGTGCTTCAGGGATCATAT 1532  
 SQE1 CTTCCCCAAAACGAATCTGGCTTGGAGCCAAATTAATCTCGGGAGCATCGGGGATAATAT 1352

SQE2 TTCCAATAATTAAGCAGAGGGAGTTAGACAAATGTTCTTCCCTCGTACAATCCCTGC-C 1737  
 SQE3 TTCCAATAATAAAGCAGAAGGTGTGAGGCAAAATGTTCTTCCCTCGAACTATTCTGC-C 1591  
 SQE1 TTCCGATAATAAAGCGGAAGGAGTTAGGCAGATGTTTTTCCAGCAACTGTACCTGCAT 1412

SQE2 CTGAAAACACGGGTTGCGCCTCAAGTACCAACCAAGGTCCGTGAAGCATTCATCACCGCT 1140  
 SQE3 CTCAAAACAATGGTTGCACCGCAGGTACCACCTCAGATCCGTGATGCTTTCATTTCTGCA 994  
 SQE1 TTGAAGACTGTTGTGGCTCCTCAGATGCCTCATGAGGTCTATGACTCTTTCATTGCTGCG 814

SQE2 GTTGAGAAAGGTAATATCAGAACCATGCCAAACCGAAGCATGCCAGCTGATCCGATTCT 1200  
 SQE3 GTCGAAAAAGGTAACATAAGAAACAATGCCGAACAGAAGCATGCCAGCTGATCCAATTCAT 1054  
 SQE1 GTTGATAAAGGAAATATTAAGTCCATGCCAAACAGAAGCATGCCAGCTTCTCCTTATCT 874

SQE2 ACTCCTGGAGCTCTTCTTCTTGGTGATGCATTCAACATGAGACATCCTTTAACCGTGGT 1260  
 SQE3 ACACCTGGAGCTTTGCTTTTAGGTGATGCGTTCAACA **TGCGCCATCCTCTTACTGGA**GGT 1114  
 SQE1 ACTCCAGGGGCTCTGTTAATGGAGATGCATTTAACATGCGTCATCCTTTGACGGGTGGA 934

SQE2 GGGATGACCGTTGCATTGGCGGATATAGTTGTACTCCGTGATCTTCTAAGGCCAATTCGC 1320  
 SQE3 GGTATGACCGTTGCATTGTCTGATATAGTCATCCTCCGTGATCTATTGAACCCGCTCGTC 1174  
 SQE1 GGAATGACGGTTGCATTAGCTGACATTGTTGTCTGCGTAATCTCCTTAGACCGTGCGT 994

SQE2 AACCTTAATGACAAAAG--AAGCTTTGTCTAAGTATATTGAATCCTTTTACACACTACGAA 1378  
 SQE3 GACTTAACCAACAAAAG--AGTCCTTATCCAAATACATAGAATCATTTACACATTGCGGA 1232  
 SQE1 GATCTTAGTGACGGCGCTAGTCTCTG--CAAATATCTTGAATCATTTTACACTCTGCGAA 1052

SQE2 AACCTGTAGCTTCCACCATTAATACATTGGCGGATGCGTTGTATAAGGTCTTTTTAGCAT 1438  
 SQE3 AACCGGTTGCTTCGACTATCAATACACTCGCAGGCGCTTTTACAAAGTCTTTTTAGCAT 1292  
 SQE1 AGCCAGTTGCAGCAACAATCAACACCCTTGCGAATGCTCTTTACCAAGTGTTCTGTTTCAT 1112

SQE2 CTTTCAGATGAAGCAAGAACGGAATGCGTGAAGCTTGCTTCGACTATCTTAGCCTTGGAG 1498  
 SQE3 CTCCAGACGATGCAAGAAGCGAAATGCGTCGAGCTTGCTTCGATTATCTTAGCCTCGGAG 1352  
 SQE1 CAGAAAAATGAAGCAAGAAACGAGATGAGGGAAGCTTGCTTCGATTATCTGGGCCCTCGGG 1172

SQE2 GTGTTTTCTCATCTGGTCCAGTTGCATTGCTCTCTGGTTTAAACCC **TCGTCTCTGAGT** 1558  
 SQE3 GGGTTTGCTCATCAGGACCTGTGGCTTGTCTATCTGGTTTGAACCCACGACCTATGAGCC 1412  
 SQE1 GTATGTGACAAGTGACCAAGTATCTTTGCTTTCGGGTTT **GAACCCCTGACCATTAACAC** 1232

SQE2 **TAGTTCTCCAC**TTCTTTGCTGTGGCGATCTACGCTGTTTGTGCTTTAATGCTACCATTC 1618  
 SQE3 TTGTTCTTCATTTCTTCGCAGTTGCGATTTTCGGGGTTGGTCGTTTGCTTGTACCTCTCC 1472  
 SQE1 **TGTC**TGCCATTTCTTTGCGGTTGCGGTTTATGGAGTCATACGGTTGTTAATCCCATTC 1292

SQE2 CTTTCGATTGAGAGCTTTTGGCTTGGAGCTAGGATAATCTCGAGTGCTTCAAGCATCATCT 1678  
 SQE3 CGTCCGTTAAACGGTTATGGCTTGGAGCTAGACTAATCTCGAGTGCTTCAGGGATCATAT 1532  
 SQE1 CTTCCCCAAAACGAATCTGGCTTGGAGCCAAATTAATCTCGGGAGCATCGGGGATAATAT 1352

SQE2 TTCCAATAATTAAGCAGAGGGAGTTAGACAAATGTTCTTCCCTCGTACAATCCCTGC-C 1737  
 SQE3 TTCCAATAATAAAGCAGAAGGTGTGAGGCAAAATGTTCTTCCCTCGAACTATTCTGC-C 1591  
 SQE1 TTCCGATAATAAAGCGGAAGGAGTTAGGCAGATGTTTTTCCAGCAACTGTACCTGCAT 1412

# APPENDIX VI: GENES AGI CODES AND PROBE SET ID NUMBERS INVOLVED IN THE PHYTOSTEROL BIOCHEMICAL PATHWAY IN *ARABIDOPSIS*

Genes involved in phytosterol biosynthesis in *A. thaliana*. Product, enzyme, gene abbreviation, AGI code and probe set ID information are included.

Product	Enzyme	AGI code	Probe set ID
24-Methylene cycloartenol	Sterol C-methyltransferase	At5g13710	250254_at
Cycloeucalenol	24-methylenecycloartenol	At4g12110	254860_at
Cycloeucalenol	24-methylenecycloartenol	At4g22750	254279_at
Obtusifoliol	Cycloeucalenol	At5g50375	248511_at
Delta,14 Sterol	Obtusifoliol C14 methyl	At2g17330	264877_at
Delta 14 Sterol	Obtusifoliol C14 methyl	At1g11680	262820_at
4 $\alpha$ -Methylfecosterol	C14 reductase	At3g52940	251995_at
24-Methylene Iophenol	Delta 8Delta 7 Isomerase	At1g20050	261228_at
24-Ethylidene Iophenol	Sterol C-methyltransferase	At1g76090	261727_at
24-Ethylidene Iophenol	Sterol C-methyltransferase	At1g20330	255885_at
Avenasterol	Sterol C-4 methyl oxidase	At2g29390	266289_at
Avenasterol	Sterol C-4 methyl oxidase	At1g07420	261076_at
5- Dehydroavenasterol	Delta 7-Sterol-C-5-	At3g02580	258484_at
5- Dehydroavenasterol	Delta 7-Sterol-C-5-	At3g02590	258404_at
Isocuposterol	Delta-Sterol-C-5-reductase	At1g50430	261865_at
Sitosterol	Sterol C24-reductase	At3g19820	257938_at
Stigmasterol	C-22-Sterol desaturase		

## **APPENDIX VII: PUTATIVE B-AMYRIN SYNTHASE NUCLEIC ACID SEQUENCE**

A putative  $\beta$ -amyrin synthase partial sequence was elucidated and supplied by Sean Mayes 2004 (unpublished data) from Restharrow. On the partial sequence, 3' rapid amplification of cDNA ends was carried out to elucidate an almost complete sequence, missing the first 7 amino acid residues on the 5' terminal. The sequence for an extended  $\beta$ -amyrin synthase like- nucleic acid sequence, Metclone 15:

```
TTGAATGATCCATACTTATTTAGCACAAATAAGTTTGTGGGAAGA  
CAAACATGGGAATATGATCCAGAAGCAGGTAGTGAGGAAGAGA  
GGGCTCAAGCAGAAGAAGCTCGTATCAATTTCTATAACAACCGC  
TTCGAGTTTAAACCATGCGGTGACCTCCTCTGGCGTTTTTCAGGTT  
CTAAGAGAAAATAATTTCAAACAAACACTAGATGGTGTAAGAT  
AGAAGATGGAGAAGAGATAACATATGAAAAAGCCACAACAACA  
TTGAGAAGGGGACACATCACCTAGCAGCATTGCAAACACTAGTGA  
TGGCCATTGGCCTGCTCAAATTGCTGGTCCTCTCTTTTTTCATGCCT  
CCCTTGGTTTTTTTGTGTCTACATTACTGGACATCTTGATTCTGTAT  
TCCCACAAGAGCATCGCAAAGAGATTCTCCGTTACATATACTGTC  
ACCAGAATGAAGATGGAGGTTGGGGGCTACACATTGAGGGGTCAC  
AGCACCATGTTTTGTACTGCACTCAACTATATATGCATGCGAATT  
CTAGGAGAAGGGCCCGATGGAGGTCACGACAATGCTTGTGCTAG  
AGCAAGAAAATGGATTTCGAGAACACGGTAGTGTAACGCATATAC  
CTTCCTGGGGAAAAACGTGGCTTTCTATACTTGGTCTATTTGATT  
GGTGTGGAAGCAATCCAATGCCCCCTGAGTTTTGGATCCTTCCTT  
CATTTCTTCCTATGCATCCAGCTAAAATGTGGTGTTATTGTGCGATT  
GGTATACATGCCTATGTCTTACTTGTATGGCAAGAGGTTTGTGGC  
TAGAATCACACCACTCATCTTAGAGTTGAGAGAAGAAGCTCTATA  
CTCAACCTTATGAAAAAGTTAATTGGACAAAATCACGACACCTA  
TGTGCAAAGGAAGATCTTTACTATCCTCATCCGTTGATACAAGAT  
TTGATATGGGATAGCTTATACATATTTACTGAACCACTTCTGACC  
CGTTGGCCTTTCAACAAGCTTGTGACAGACAGAAAGCTCTTCAAGTA  
ACAATGAAGCATATCCACTACGAGGATGAGAATAGTCGGTACCT  
AACCATTGGGTGTGTGGAAAAGGTTTTATGTATGCTTGCTTGTTG
```

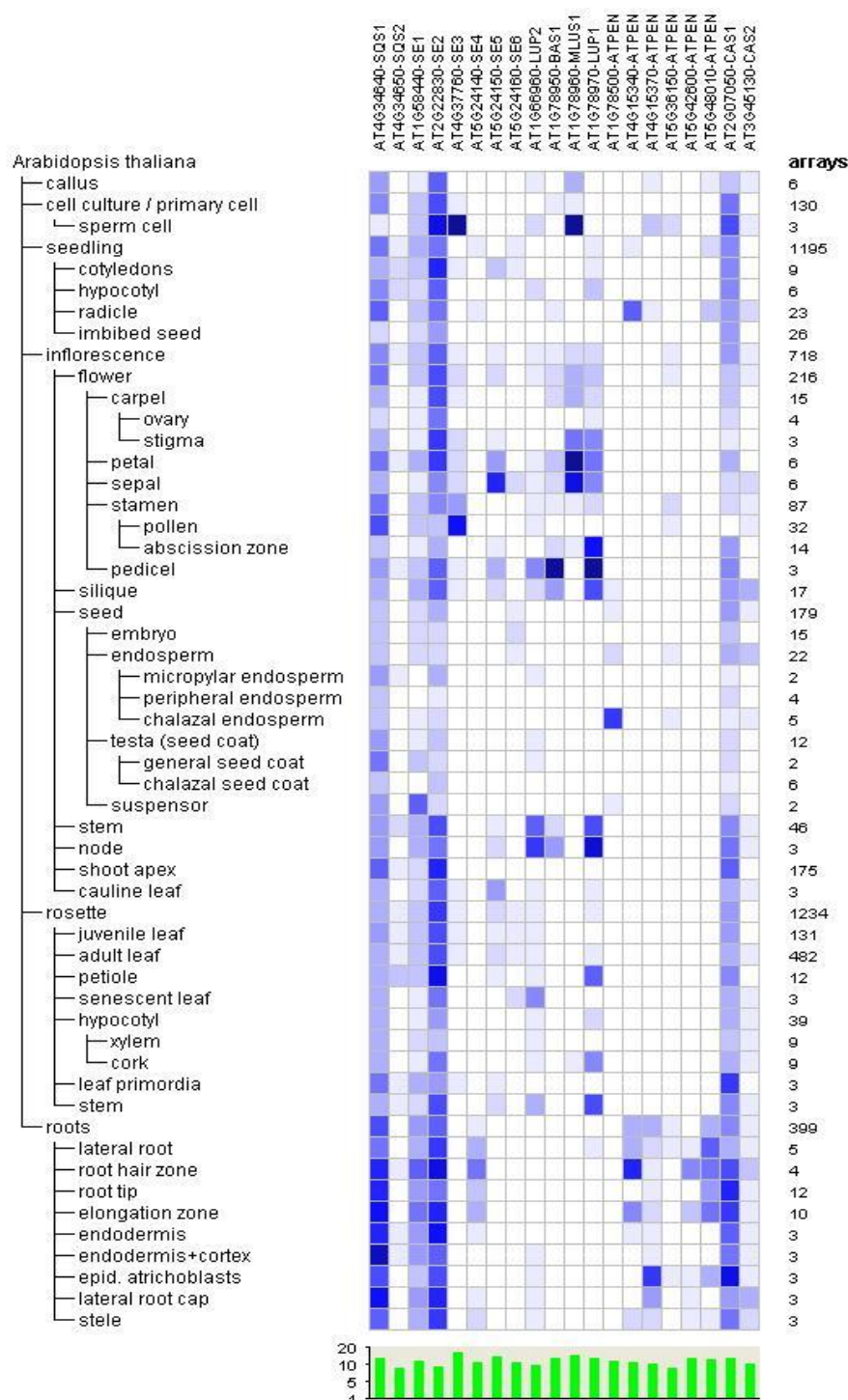
GGTGGAAGATCCAAATGGGGATGCTTTCAAGAAGCATCTTGCAA  
GGGTCCCAGATTACTTATGGATGTCAGAAGATGGAATGACAATG  
CAGAGTTTTGGTAGCCAAGAATGGGATGCTGGTTTTGCAGTTCAA  
GCTTTGCTTGCCACTAACCTAGTTGAGGAAATTGGACCTGCACTT  
GCGAAAGGACATGATTTTATCAAGAAATCTCAGGTTACAGAAAA  
CCCTTCTGGAGATTTTAAGAGTATGCATCGTCATATTTCTAAAGG  
CTCGTGGACCTTCCCCGATCAAGACCATGGATGGCAAGTTTCTGA  
TTGCACCGCTGAAGGTTTGAAGTGTTGTCTACTTTTATCGTTGTTG  
CCTCCAGAAATTGTGGGGGAAAAGATGGAACCAGAAAGGTTATA  
CGATTCAGTCAATCTCCTGTTATCACTTCAGAGTAAAGAGGGTGG  
TTTGTGTCAGCATGGGAACCAGCAGGAGCTCAAGAGTGGTTAGAAC  
TACTCAATCCCACAGAATTTTTTGCGGACATTGTAGTTGAACATG  
AATATGTAGAGTGCACTGGATCAGCAATTCAAGCTTTAGTTTTAT  
TCAAGAAGTTATATCCAGGGCATAGAAAGAAAGAAATAGAGAAT  
TTCATTGCCAATGCAGTTCGATTCCTTGAAGATACACAAACAGCT  
GATGGTTCATGGTATGGAACTGGGGAGTTTGCTTCACTTATGGT  
TCTTGGTTTGCTCTTGGTGGTTTAGCAGCTGTTGGCAAGACTTAT  
GGCAATTGCGCTGCTATTCGCAAAGCTGTTAAGTTTCTTCTCACC  
ACACAGAGAGAGGATGGTGGGTGGGGGGAGAGCTATCTTTCAAG  
CCCAAAAAGATATATGTACCGCTCGAAGGGAGCCGGTCAAATG  
TTGTACATACTGCATGGGCTCTTATGGGTTTAATTCATGCTGGCC  
AGGAAGAGAGAGACCCTACTCCTCTTCATCGTGCTGCAAAATTG  
CTCATTAATTCCCAATTAAAAGAAGGCGATTGGCCCCAAGAGGA  
AATCACAGGAGTATTCATGAAAACTGTATGTTGCATTACCCAAT  
GTATAGAGATATTTACCCATTGTGGGCTCTAGCAGAGTATCGAA

Amino acid alignment of putative  $\beta$ -amyrin synthase from Restharrow with  
*P. sativum* and *G. Glabra*,  $\beta$ -amyrin synthase and multifunctional synthase  
from *P. sativum*

ONONIS-Putative PSATIVUM-BAS GGLABRA-BAS PSATIVUM-Multi	MWRLKIAEGGNDPYLFSTNNFVGRQTWEYDPEAGSEEEERAQAEERARINFYNNRFEVKPCG 60 MWRLKIAEGGNDPYLFSTNNFVGRQTWEYDPEAGSEEEERAQVEEARRNFYNNRFEVKPCG 60 MWRLKIAEGGKDPYIYSTNNFVGRQTWEYDPPGGTPEERAQVDAARLHFYNNRFQVKPCG 60 MWKLKIGDGGKDRNIFSTNNFVGRQTWEFDPDAGTSQEKAQVEAARQHIFYDNRFVKAACS 60
ONONIS-Putative PSATIVUM-BAS GGLABRA-BAS PSATIVUM-Multi	DLLWRFQVLRENNFKQTLGDKIEDGEEITYEKATTTLLRGTHHLAALQTSBGHWPAQIA 120 DLLWRFQVLRENNFKQTIIGGVKIEDEEITYEKTTTTLLRGTHHLATLQTSBGHWPAQIA 120 DLLWRFQILRENNFKQTIASVKIGDGEITYEKATTAVRRAAHHLALQTSBGHWPAQIA 120 DLLWRFQILKEKNFKQTIESVKIKDEEIESEENVAITLRAVHHLSTLQSNDBGHWPALNA 120
ONONIS-Putative PSATIVUM-BAS GGLABRA-BAS PSATIVUM-Multi	GPLFFMPPLVFCVYITGHLDSVFPQHRKEILRYIYCHQNEGGWGLHIEGHSTMFCTAL 180 GPLFFMPPLVFCVYITGHLDSVFPPEHRKEILRYIYCHQNEGGWGLHIEGHSTMFCTAL 180 GPLFFLPPLVFCMYITGHLDSVFPPEYRKEILRYIYCHQNEGGWGLHIEGHSTMFCTAL 180 GPLFYFPPLVFCMYVTGHLDSIFPYEYRKEILRYIYCHQNEGGWGLHVEGHSIMFCTVL 180
ONONIS-Putative PSATIVUM-BAS GGLABRA-BAS PSATIVUM-Multi	NYICMRILGEGPDGGHDNACARARKWIREHGSVTHIPSWGKTWLSILGLFDWCGSNPMPP 240 NYICMRILGEGPDGGEDNACVRARNWIRQHGGVTHIPSWGKTWLSILGVFDWLGSNPMPP 240 NYICMRILGEGPDGGQDNACARARKWIHDHGGVTHIPSWGKTWLSILGVFDWCGSNPMPP 240 NYICMRILGEGPNGGKEDACARARKWIHDHGSVTHVSSWGKTWLSVLGIFDWCASNPMPP 240
ONONIS-Putative PSATIVUM-BAS GGLABRA-BAS PSATIVUM-Multi	EFWILPSFLPMHPAKMWCYCRLVYMPMSYLYGKRKFVARITPLILELREELYTPQYKVNW 300 EFWILPSFLPMHPAKMWCYCRLVYMPMSYLYGKRKFVGPITPLILQLREELHTEPYEKINW 300 EFWILPSFLPMHPAKMWCYCRLVYMPMSYLYGKRKFVGPITPLILQLREELTEPYEKVNW 300 EFWMLPSFLLKHPAKMLCYCRLVYMPMSYLYGKRKFVGPITPLILMLREELLTPQYKVNW 300
ONONIS-Putative PSATIVUM-BAS GGLABRA-BAS PSATIVUM-Multi	TKSRHLCAKEDLYYPHPLIQDLIWDLSYIFTEPLLTRWPFNKLVRQKALQVTMKHIHYED 360 TKTRHLCAKEDIYYPHPLIQDLIWDLSYIFTEPLLTRWPFNKLVRKRALEVTKMHIHYED 360 KKARHQCACEDLYYPHPLIQDLIWDLSYIFTEPLLTRWPFNKLVRKALQVTMKHIHYED 360 KKTRHLCAKEDLYYPHPLIQDLIWDLSYIFVEPLLTHWPFNKLRLKALQVTMKHIHYED 360
ONONIS-Putative PSATIVUM-BAS GGLABRA-BAS PSATIVUM-Multi	ENSRYLITIGCVEKVLCLACWVEDPNGDAFKKHLARVPDYLWMSDGMQMSEFGSQEWDA 420 ENSRYLITIGCVEKVLCLACWVEDPNGDAFKKHARVPDYLLWIS EDGMTMQSFGSQEWDA 420 ETSRYLITIGCVEKVLCLACWVEDPNGDAFKKHLARVPDYLVWVSDGMTMQSFGSQEWDA 420 ENSRYLITIGCVEKVLCLACWVEDPNGDAFKKHLARLPDYLVWVSDGMTLHSEFGSQTWDA 420
ONONIS-Putative PSATIVUM-BAS GGLABRA-BAS PSATIVUM-Multi	GFAVQALLATNLVEEIKPALAKGHDFFIKKSQVTENPSGDFKSMHRHISKGSWTFPDQDHG 480 GFAVQALLATNLIEEIKPALAKGHDFFIKKSQVTENPSGDFKSMHRHISKGSWTFSDQDHG 480 GFAVQALLATNLVEEIAPTLAKGHDFFIKKSQVRDNPBGDFKSMYRHISKGSWTFSDQDHG 480 SLIIQALLATNLIEDVGPIITKAHEFIKKSQVRDNPBGDFKSMYRHISKGSWTFSDQDHG 480
ONONIS-Putative PSATIVUM-BAS GGLABRA-BAS PSATIVUM-Multi	WQVSDCTAEGCLKCCLLSLLPPEIVGEKMEPERLYDSVNLLLSLQSKGGLSAWEPAGAQ 540 WQVSDCTAEGCLKCCLLSLLPPEIVGEKMEPERLFDVNLLLSLQSKKGGLAAWEPAGAQ 540 WQVSDCTAEGCLKCCLLSMLPPEIVGEKMEPERLYDSVNLLLSLQSKKGGLSAWEPAGAQ 540 WQVSDCTAESLKCCLLSMLPPEIVGEKMEPEMLYDSVNLLLSLQGGKGLPAWEPSEAV 540
ONONIS-Putative PSATIVUM-BAS GGLABRA-BAS PSATIVUM-Multi	EWLELLNPTEFFADIVVEHEYVECTGSAIQALVLFKKLYPGHRKKEIENFIANAVRFLED 600 EWLELLNPTEFFADIVVEHEYVECTGSAIQALVLFKKLYPGHRKKEIENFIANAVRFLED 600 EWLELLNPTEFFADIVVEHEYVECTGSAIQALVLFKKLYPGHRKKEIENFIANAVRFLED 600 EWLELNPTEFFLEIVVEREYVECTGSAIQALVLFKKLYPEHRKKEVENFIANAVRFLEY 600
ONONIS-Putative PSATIVUM-BAS GGLABRA-BAS PSATIVUM-Multi	TQTADGSWYGNWGVCFYGSWFALGGLAAVGKTYGNCAAIRKAVKFLTTQREDGGWGES 660 TQTEDGSWYGNWGVCFYGSWFALGGLAAAGKTYTNCAAIRKVFLTTQREDGGWGES 660 TQTADGSWYGNWGVCFYGSWFALGGLAAAGKTFANCAAIRKAVKFLTTQREDGGWGES 660 KQTSDGSWYGNWGICFTYGSWFALNGLVAAGKTYDNCAAIRKGVFLTTQREDGGWGES 660
ONONIS-Putative PSATIVUM-BAS GGLABRA-BAS PSATIVUM-Multi	YLSSPKKIYVPLEGRSNVHTAWALMGLIHAGQSERDPTPLHRAAKLLINSQLEGGDWP 720 YLSSPKKIYVPLEGRSNVHTAWALMGLIHAGQSERDPTPLHRAAKLLINSQLEGGDWP 720 YLSSPKKIYVPLEGRSNVHTAWALMGLIHAGQERDAPLHRAAKLLINSQLEGGDWP 720 HLSSSKKIYVPLERSQSNIVQTSWAIMGLIHAGQMERDPTPLHRAAKLLINSQLEGGDWP 720
ONONIS-Putative PSATIVUM-BAS GGLABRA-BAS PSATIVUM-Multi	QEEITGVFMKNCMLHYPMYRDIYPLWALAEYRRRVPLP----- 758 QEEITGVFMKNCMLHYPMYRDIYPLWALAEYRRRVPLP----- 758 QEEITGVFMKNCMLHYPMYRDIYPMWALAEYRRRVPLPSTPCLT 765 QEEITGVFMKNCMLQYAMYRDIYPTWALAEYRRRILLASPAVAT- 764

The ClustalW2 (EBI), multiple amino acid alignment tool was used to align  $\beta$ -amyrin synthases.  $\beta$ -amyrin synthase (BAA97558) in *Pisum sativum*,  $\beta$ -amyrin synthase (BAA89815) in *Glycyrrhiza glabra*, a multifunctional Oxidosqualene cyclase (PSM: BAA97559) from *P. sativum* and an Oxidosqualene cyclase (*O. spinosa*) were aligned at the translated amino acid level.

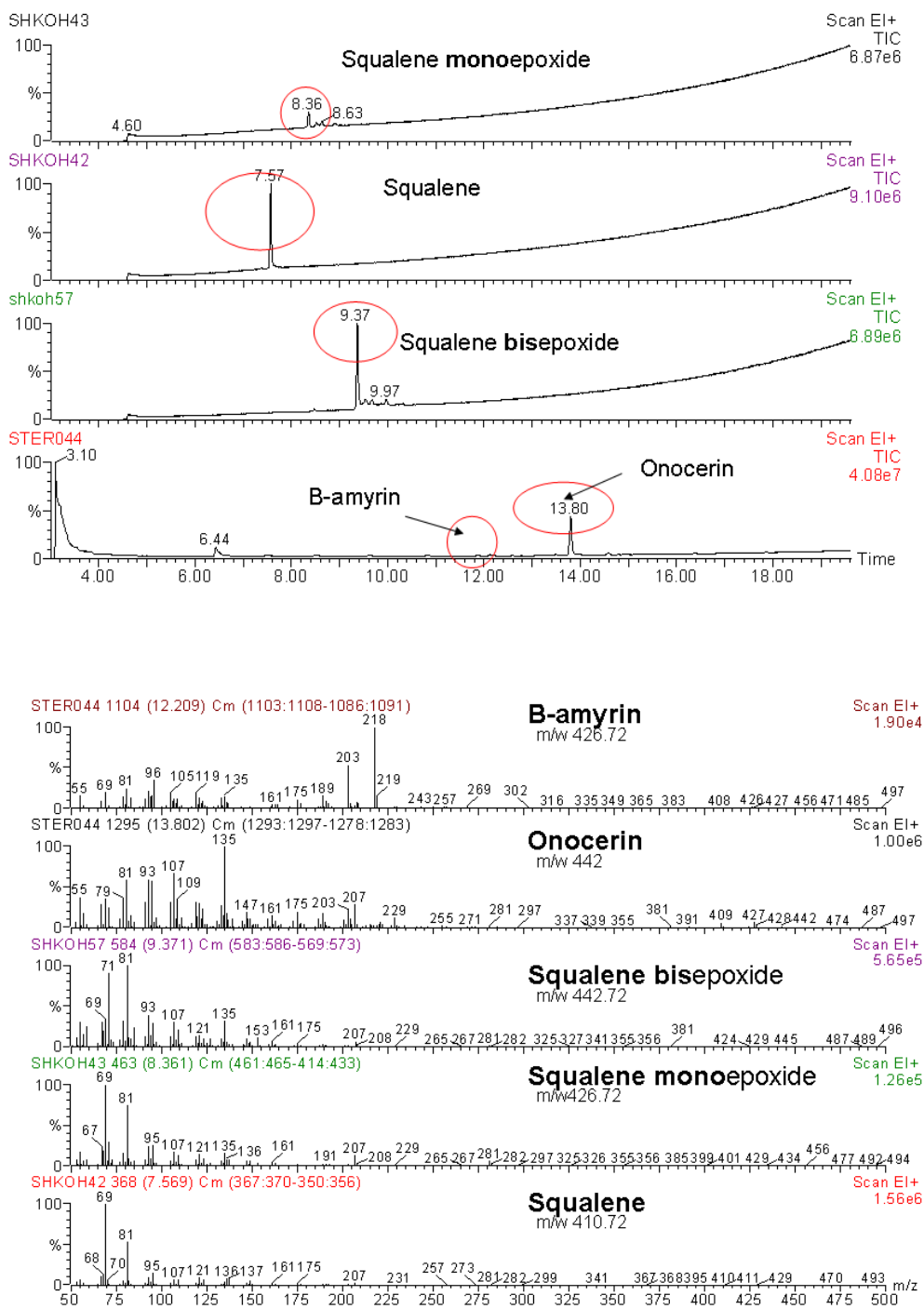
# APPENDIX VIII: GENE EXPRESSION PROFILES; META-PROFILE ANALYSIS TOOL FROM GENEVESTIGATOR



A detailed tissue specific gene expression profiles from the Meta-profile analysis tool from GENEVESTIGATOR. The heat map shows mean gene expression values within individual tissues. The histogram adjacent to the heat map indicates the number of experiments used to generate the mean signal value.

## APPENDIX IX: GC/EIMS STANDARDS: GC, MS AND FRAGMENTATION IONS LISTED

The molecular weight and formula were used to identify the likely retention time (GC) and key molecular ion (MS). The above standards were derivitised using a Hexane/diethyl ether NSL polar phase extraction and GC/MS:



Derivatised standards showing retention times on Mass Spectra; Squalene, Squalene monoepoxide, Squalene bisepoxide  $\beta$ -amyrin and  $\alpha$ -onocerin and all terpenoids.

<b>Alkane</b>	<b>GC-EIMS (relative intensity)</b>
Hexacosane $C_{26}H_{54}$	<i>m/z</i> : <b>366</b> (10) 365 (19), 351 (5), 337 (4), 323 (4) 309 (5) 295 (4) 281 (4) 267 (5) 253 (6) 239 (6), 225 (6), 211 (6), 197 (6), 183 (7), 169 (7), 155 (8), 141 (9), 127 (12), 113 (17), 111 (18) 99 (22), 97 (28), 85 (70), 71 (95), 57 (100).
Octacosane $C_{28}H_{58}$	<i>m/z</i> <b>394</b> (18), 379 (6), 365 (4), 337 (6), 323 (5), 309 (5), 295 (5), 281 (6), 267 (5), 253 (5), 239 (7), 225(6), 211 (7), 197 (8), 183 (6), 169 (8), 155 (7), 141 (9), 127 (12), 113 (15), 111 (25), 99 (26), 97 (39), 85 (72), 71 (90), 57 (100).
Triacontane $C_{30}H_{62}$	<i>m/z</i> : <b>422</b> (15), 393 (5), 779 (6), 365 (6), 351 (7), 337 (7), 323 (6) 309 (6) 295 (6) 281 (7) 267 (6) 253 (6) 239 (9), 225 (7), 211 (9), 197 (8), 183 (10), 169 (8), 155 (10), 141 (13), 127 (15), 113 (20), 111 (27), 99 (26), 97 (35), 85 (80), 71 (95), 57 (100).
Onocerin	GC-EIMS (relative intensity) <i>m/z</i> : <b>442</b> (5), 427 (10), 424 (5), 413 (5), 409 (6) 391 (4), 229 (15), 203 (22), 189 (22 ), 175 (20 ), 161 (15), 147 (30), 135 (100), 123 (35), 107 (50), 95 (90), 81 (75), 69 (50), 53 (40).
$\beta$ -amyrin	<i>m/z</i> 426(4),411(3), 257(3), 218 (100), 203 (50), 189 (17), 135 (18), 119 (18) 95 (18), 69 (22), 55(12).
Sitosterol	<i>m/z</i> <b>414</b> (70), 396 (100), 381 (14), 354 (4), 329 (18), 303 (18), 255 (18), 213 (15), 199 (5), 173 (16), 145 (20), 121 (16), 107 (27) 95 (28) 81 (25), 69 (23), 53 (3).

<b>Standard</b>	<b>Formula</b>	<b>CAS</b>	<b>Molecular</b>
Triacontane	$C_{30}H_{62}$	638-68-6	422
Octacosane	$C_{28}H_{58}$	630-02-4	394
Squalene	$C_{30}H_{50}$	111-02-4	410.72
2,3-oxidosqualene	$C_{30}H_{50}O$	54274-53-2	426.72
2,3;22,23-dioxidosqualene	$C_{30}H_{50}O_2$	77008-90-3	442.72
$\beta$ -amyrin	$C_{30}H_{50}O$	559-70-6	426.72
$\alpha$ -onocerin	$C_{30}H_{50}O_2$	511-01-3	442

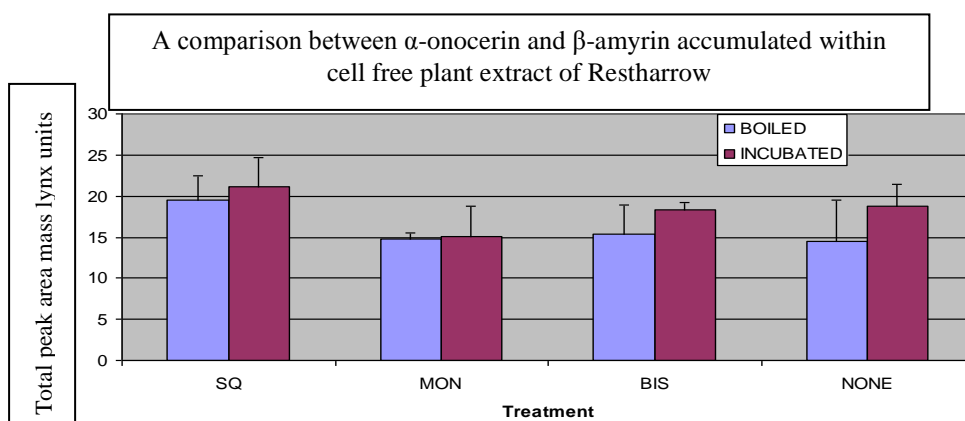


A summary table of the NSL content within roots within Restharrows and liquorice as a non-species control. The concentration of metabolite along with the standard deviation is shown.

NSL	<i>O. maritima</i>		<i>O. repens</i>		<i>O. rotundifolia</i>		<i>O. spinosa</i>		<i>G. glabra</i>		<i>O. pusilla</i>	
	Mg/g	StDev	Mg/g	StDev	Mg/g	StDev	Mg/g	StDev	Mg/g	StDev	Mg/g	StDev
<b>Stigmasterol</b>	0.099	0.020	0.037	0.011	0.009	0.002	0.141	0.039	0.198	0.091	0.191	0.004
<b>Bis-epoxide</b>	0.156	0.001	0.063	0.0066	0.000	0.000	0.250	0.069	0.000	0.000	0.000	0.000
<b>Sitosterol</b>	0.249	0.014	0.216	0.047	0.100	0.027	0.514	0.029	0.344	0.020	0.275	0.082
<b>B-amyrin</b>	0.662	0.001	0.182	0.092	0.011	0.011	2.494	0.109	0.091	0.070	0.079	0.021
<b>Mono-epoxide</b>	0.049	0.006	0.009	0.000	0.000	0.000	0.048	0.001	0.031	0.027	0.000	0.000
<b>Onocerin</b>	3.732	0.000	1.937	0.064	0.000	0.000	4.651	0.184	0.000	0.000	0.000	0.000

A summary table of the NSL content within leaf within Restharrows and liquorice as a non-species control. The concentration of metabolite along with the standard deviation is shown.

NSL	<i>O. maritima</i>		<i>O. repens</i>		<i>O. rotundifolia</i>		<i>O. spinosa</i>		<i>G. glabra</i>		<i>O. pusilla</i>	
	Mg/g	StDev	Mg/g	StDev	Mg/g	StDev	Mg/g	StDev	Mg/g	StDev	Mg/g	StDev
<b>Stigmasterol</b>	0.038	0.030	0.023	0.009	0.009	0.002	0.086	0.013	0.099	0.039	0.053	0.009
<b>bis-epoxide</b>	0.000	0.000	0.000	0.000	0.000	0.000	0.800	0.067	0.000	0.000	0.000	0.000
<b>Sitosterol</b>	0.248	0.123	0.203	0.101	0.131	0.028	2.494	0.109	0.574	0.020	0.823	0.020
<b>B-amyrin</b>	0.007	0.003	0.004	0.003	0.116	0.017	0.012	0.002	0.147	0.080	0.498	0.018
<b>Mono-epoxide</b>	0.016	0.008	0.005	0.000	0.035	0.009	0.027	0.000	0.091	0.039	0.067	0.021
<b>Onocerin</b>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000



**Figure 5.41** A comparison between  $\alpha$ -onocerin and  $\beta$ -amyrin within incubated and boiled samples. Substrates added; squalene (SQ), 2,3-oxidosqualene (MONO), 2,3;22,23-dioxidosqualene (BIS).